ADMISSIBLE AND MINIMAX LINEAR ESTIMATION THEORY

by

PEDRO ALSON HARAN

Thesis

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY AT THE

UNIVERSITY OF LONDON

DEPARTMENT OF MATHEMATICS

IMPERIAL COLLEGE

ABSTRACT

This thesis studies admissibility, in the context of linear model theory, under general quadratic risks, when only linear estimators are considered. Algebraic representations of admissible linear estimators are obtained for the following cases: the space of the mean parameter is $\mathbb{R}^p$ and the design matrix X is full rank or non full rank; the space of the mean parameter is a subspace of $\mathbb{R}^p$ and X is full rank. Minimax properties of linear estimators over special cubes and any ellipsoid centred at the origin are given. In particular the following questions are answered: given a constraint region of such a kind and a quadratic risk what is the minimax linear estimator under this risk for this region? Given a linear estimator and a quadratic risk on what regions the estimator is linear minimax under the quadratic risk. The region where an admissible linear estimator has better matrix quadratic risk than the best linear unbiased estimator is characterized. An optimal property of restricted best linear estimators is proved and finally admissible linear estimators are characterized in terms of minimax properties.

ACKNOWLEDGEMENTS

I am very grateful to my supervisor, Dr. H.P. Wynn, for his constant academic (and non-academic) support.

The existence of a real school (of statistics) at Imperial College made my permanence there more enjoyable and fruitful. This school is mainly the creation of Professor D.R. Cox and it is for me a pleasure to thank him for it.

I would also like to thank the staff of the Mathematics Library, for their help, Mrs. Robertson for her excellent typing and Mrs. P.A. Easton for her assistance on a number of occasions.

Without the financial support of Conicit, Caracas, Venezuela, it would have been impossible for me to realize this work.

Finally, something important would be missing in this acknowledgement if I did not thank Irene for her presence beside me, in those long years outside my country and for many other things.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

To set the background for this thesis it is necessary to sketch some of the main branches in the evolution of the theory of statistical linear models. We shall not attempt to give a full history since the introduction of least squares by Gauss (see the 1855 book) but rather make use of more recent work, some of which themselves contain extensive bibliographies.

Before the 1960's the theory separates roughly into two main groups. The first tries to interpret many of the statistical procedures and concepts arising in practice in terms of least squares. Some of the questions are: the definition of the best linear unbiased estimator (BLUE) as a least squares estimator (LSE) when the variance of the model is of the form $\sigma^2 V$ (Aitken, 1935), when X is a non full rank matrix (Bose 1944, Rao 1946) or when the parameter of the mean is restricted to lie in a subspace (Rao 1945), the formulation of hypothesis testing in terms of least squares (Rao 1946). The second group, with a smaller literature (Durbin and Kendall 1951), but nevertheless important, develops geometrical understanding of the least squares estimator. It becomes clear that a least squares estimator can always be seen as the projection (using an appropriate inner product) over some subspace contained in the range of the design matrix X.

Kruskal's free coordinate approach (1961) extends existing geometrical interpretations to other practical situations. The relation between BLUE and LSE is studied again (Zyskind 1967, Kruskal 1968).

The g-inverses are adopted as a standard algebraic device to solve problems related to the estimation of the mean. This work has been extensive since the early 1960's (see Mitra and Rao 1971 for a good account). By the end of the 1960's the theory of BLU estimation had been thoroughly worked and indeed it was probably a common belief that BLUE's were the only useful linear estimators.

The beginning of the 1970's brought fresh life into the topic. Mainly motivated by computational problems, Hoerl and Kennard (1970a, 1970b) used perturbations of the LSE of the form $M_k Y = (X'X + kI)^{-1}X'Y$ as estimators. They were called ridge estimators. It was a turning point for the estimation procedures of the mean of the linear model: the new estimators had interesting statistical properties which qualified them as possible competitors of the BLUE's. A wide field was open up to research. As a first consequence, a host of new estimators, generalizations of the ridge estimator, became available. Their increasing use in practical situations generated a considerable number of theoretical and applied studies. On the applied side the paper by Gotô (1979) (see also its bibliography) is a guide to the developments in the area. Three main avenues of theoretical research can be distinguished.

Efron and Morris (1973) developed the ideas of Stein (1956) whose famous work showed the inadmissibility of the ordinary unbiased estimator of the multivariate normal mean and hence the LSE in regression for dimensions greater than two. Exploiting some of the connections between shrinkage estimators and ridge estimators Thisted (1976) gave a "ridge rule" which had uniformly smaller risk than the relevant BLUE (in the normal case). There has been considerable subsequent work in this area.

Although with the assumption of normality a wide range of practical situations can be covered it is also important to have properties which depend only on second order assumptions. Kuks and Olman 1972, Kuks (1972) and later Bunke (1975), Lauter (1975) studied minimax properties of those estimators in the set of linear estimators, which depend only on second order assumptions.

Finally another important contribution was to recognize that all those new estimators had a common property: the class of the linear admissible estimators was defined. This gradually became apparent through the works of Cohen (1966), Shinozaki (1975) and mainly Rao (1976) who gave a clear characterization of them. (See Definition 2.1 and Corollary 2.3.1 in this thesis.) Hoffmann (1977) studies the class of admissible linear estimators when the mean is restricted to be in an ellipsoid centred at the origin.

In the meantime the theory of BLUE's has produced important developments. Seely (1970a,b) emphasizes the notion of estimable function due to Bose (1944) to build up an elegant and flexible theory, and applies the results to the estimation of the components of variances in the mixed linear model. Eaton (1970) gives a free coordinate approach of the estimation of the mean in a mixed linear model. Gnot et al (1980) is the most refined prolongation of this work, their list of references gives a good idea of the main intermediary developments. Rao (1971) also builds up a general theory of BLUE's. Its main concern is to reduce most of the problems of BLU estimation and testability to the one of calculating a g-inverse.

Despite all the contributions mentioned above a close look at the subject reveals a number of gaps and shortcomings. It is the main motivation for this thesis to try and make up for some of these.

Here is a list:

(i) Some useful properties have been proved only for particular kinds of admissible linear estimators. An immediate question is to see if those properties are in fact a property of all the admissible linear estimators. For example: Farebrother (1976) and Obenchain (1978) studied the good region for ridge and generalized ridge estimators. What is the good region for an admissible linear estimator? (see Section 5.1.). Kuks and Olman (1972) proved a minimax property for a dense subset of the set of admissible linear estimators. In Section 5.3 we extend that property.

(ii) Some of the theoretical work such as that developed by Kuks (1972) and Lauter (1975) is difficult to apply in practice because of the complexity of the algebraic formulation. It is reasonable to expect that further work in this direction would be helpful; see Chapter 4.

(iii) There is a big division between the full rank case and the non full rank case. The g-inverse algebraic machinery has reached a high degree of sophistication and is successfully applied to treat the non full rank case in the least square theory of estimation. It seems, however, that a similar application of g-inverses to admissible linear estimators in the non full rank case is more intractable (see Sections 3.2 and 3.3).

(iv) Another desirable feature, which is absent from admissible linear estimation theory is a good geometrical interpretation. This is despite the fact that the use in practice of admissible linear estimators, is restricted to those for which a clear geometrical understanding of the bias function is available (see Corollary 2.2.1 and Corollary 2.3.2).

(v) For an arbitrary admissible linear estimator only one algebraic representation has been given (see Rao 1976) but it does not seem to have attracted much attention. It can be proved that any restricted BLUE is an admissible linear estimator but the representation given by Rao does not help us to see this easily. Moreover, whereas in this representation, all admissible linear estimators look very much the same, the usual representations of subclasses of admissible linear estimators look very different (see Sections 2.1 and 2.3).

(vi) A more remarkable fact is that the concept of admissible linear estimator has been studied for the case when quadratic restrictions are imposed on the parameter (see Hoffmann, 1977) but for the intuitively easier case where only linear restrictions have to be satisfied (i.e. the parameter is restricted to be in a subspace) no work exists on admissible linear estimators (see Section 3.4).

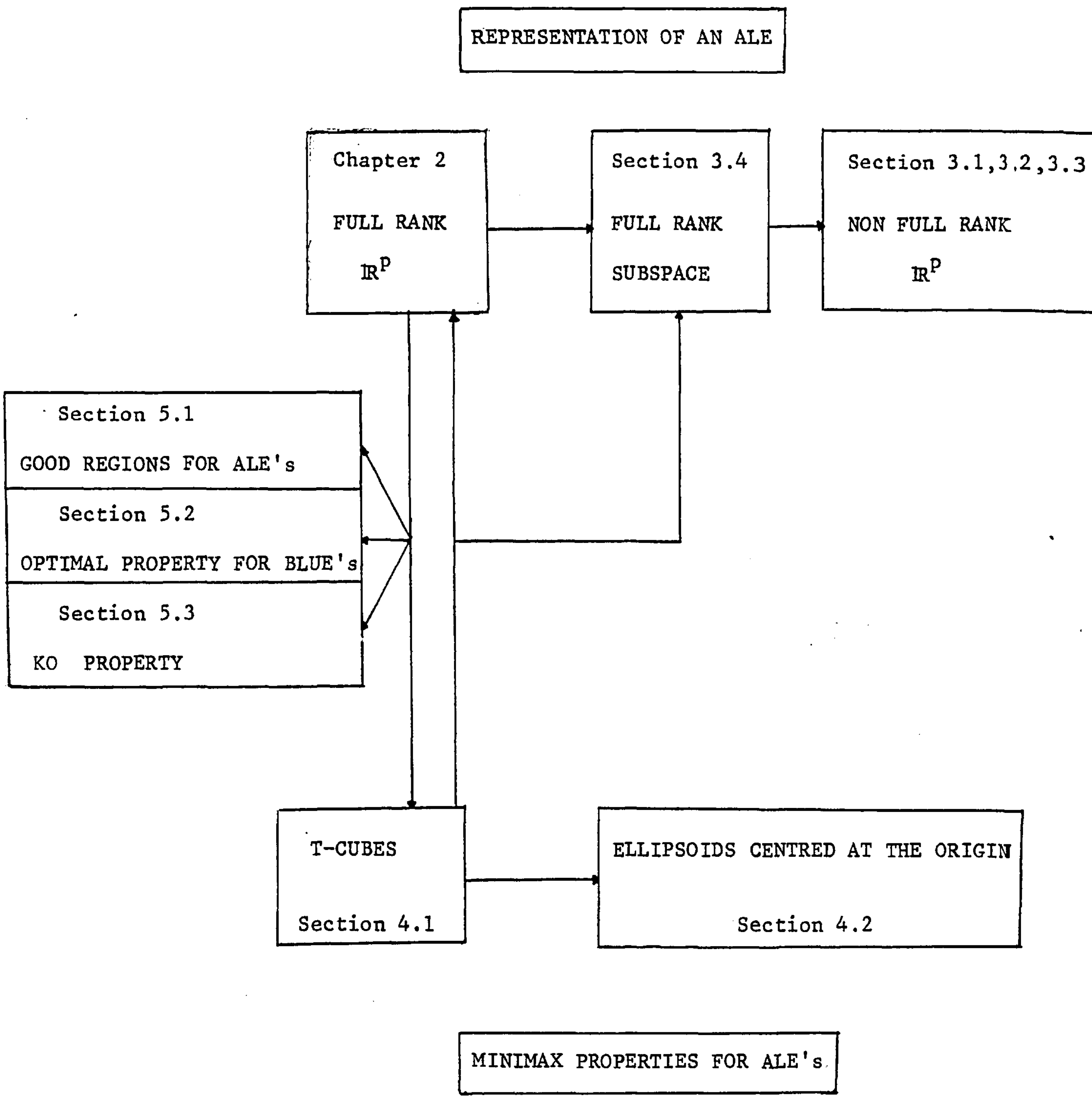
All these apparently divergent points have a common root. The aim of this thesis is to expose this. To do this a general representation of admissible linear estimators is given (Theorem 2.3). This representation is, from a purely algebraic point of view, just a step further than the one given by Rao. It is nevertheless this small step which will allow us to discover most of the relations underlying different approaches in the literature. As a byproduct of this, the set of admissible linear estimators will appear as a natural extension, from a statistical point of view, of the set of BLUE's (see end of Section 2.2).

In previous works on BLU estimation, geometrical interpretation was mainly used to give proofs of already known propositions or to understand in geometrical terms some of the algebraic procedures used. Our work goes further in that as a consequence of it we make some progress in stating and solving unsolved problems in the theory, also new algebraic procedures for old problems are suggested (see the method at the end of Section 3.2). In our approach geometry plays an essential part. The results are a natural consequence of working in the geometry defined by the inner product given by $X'V^{-1}X$ (the information matrix of the model). Algebra plays the other part precisely because it provides the notation and operational rules which gives the approach its flexibility and utility in practical situations. The main tool which enables us to take advantage of both the geometry and the algebra underlying the subject is the component estimator (see Definition 2.3).

For completeness we should mention areas not covered in this thesis but to which the work may eventually have applications. The main areas are: (1) dynamic linear system theory, much of the development of which has been in control theory. (2) Robust regression techniques introduced largely to cope with non-normal errors. (3) Special study of particular design situations (restrictions on X). (4) Computational aspects of the new theory developed.

Before giving a chart which indicates some of the main interrelations between chapters and describes the structure of the thesis, some notation will be given.

The standard notation used in set theory will be adopted. If X is a matrix, X' will denote its transpose; if C is a subset of the domain of X, X(C) will denote the image of C under X, N(X) the kernel of X and Range (X) the range of X. $\mathbb{R}$, $\mathbb{R}^p$ and $\mathcal{N}$ will denote the set of real numbers, the usual euclidean vector space and the set of natural numbers respectively. ϕ will denote the empty set and [a,b] the closed interval with extremes a and b. C-Z will denote the set difference between C and Z. $|a|$, $||v||$ will be the absolute value and the euclidean norm of a and v respectively. $\overline{\text{Lim}}$ and $\underline{\text{Lim}}$ will be the upper and lower limits. $\underset{\sim}{k}$ will denote vector k as opposed to the scalar k. The letter I will be reserved to denote the identity p×p matrix.


REPRESENTATION OF AN ALE
Chapter 2
FULL RANK
$\mathbb{R}^p$
Section 3.4
FULL RANK
SUBSPACE
Section 3.1,3.2,3.3
NON FULL RANK
$\mathbb{R}^p$
Section 5.1
GOOD REGIONS FOR ALE's
Section 5.2
OPTIMAL PROPERTY FOR BLUE's
Section 5.3
KO PROPERTY
T-CUBES
Section 4.1
ELLIPSOIDS CENTRED AT THE ORIGIN
Section 4.2
MINIMAX PROPERTIES FOR ALE's

CHART

CHAPTER 2

Let X be an $\ell \times p$ matrix of rank p, V an $\ell \times \ell$ positive definite symmetric (p.d.s.) matrix and C a subset of $\mathbb{R}^p$. The triplet (X,V,C) will be called a linear model with design matrix X, variance V and parameter space C.

Let $\mathcal{Y}_C = \{Y_\beta | Y_\beta$ is a random $\ell \times 1$ vector and $\beta \in C\}$. $\mathcal{Y}_C$ is said to satisfy the linear model (X,V,C) iff

$$E[Y_\beta] = X\beta , \qquad \forall \beta \in C.$$

$$\mathrm{Var}[Y_\beta] = V , \qquad \forall \beta \in C.$$

Where $E[Y_\beta]$ and $\mathrm{Var}[Y_\beta]$ are the expectation and the covariance matrix of Y_β. The last equality implies that the variance of Y_β is functionally independent of β. This assumption will be kept throughout the thesis.

In practice a realization Y of one of the random variables Y_β is observed and β is not known. The problem is to obtain some information about β from Y. One usual way for this is the following one: Choose (independently of the observed value Y) a $p \times \ell$ matrix M. Consider the value MY as an estimate of the value of the unknown β. When such a procedure is followed, it is said that β has been estimated linearly or that MY is a linear estimate of β or even that MY is a linear estimator of β. According to this we will use MY to designate any of the random vectors MY_β with $\beta \in C$, and we will call it a linear estimator of β. The set of linear estimators of β will be denoted by L. Since nothing is lost in clarity and to avoid superfluous symbols, we will make the following conventions:

$E_\beta[M]$	will be	$E[MY_\beta]$.
$Var[M]$	will be	$Var[MY_\beta]$.
$Cov[M,N]$	will be	$Cov[MY_\beta, NY_\beta]$.

(Here Cov means covariance).

As a result of long theory and practice some elements of L are of particular interest from a statistical point of view. The next paragraphs will give a short account about those particular estimators and some properties that are important statistically. They also will help to situate the content of the chapter in a proper perspective.

A linear estimator which has been widely used in practice is the generalized least squares estimator (GLSE). One standard expression for it is:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y .$$

Another common expression for $\hat{\beta}$ is obtained using the spectral decomposition of $X'V^{-1}X$ (see Watson 1967) and it is given by:

$$\hat{\beta} = \sum_{i=1}^{p} \frac{1}{\lambda_i} v_i v_i' X'V^{-1}Y ,$$

where $\{v_i\}_{i=1}^{p}$ is a complete set of normalized eigenvectors of $X'V^{-1}X$ and λ_i are the corresponding eigenvalues. Since $\{v_i\}_{i=1}^{p}$ is a basis of $\mathbb{R}^p$, for any $\beta \in \mathbb{R}^p$ there are scalars $b_1,\ldots,b_p$ such that $\beta = \sum_{i=1}^{p} b_i v_i$. Let $\hat{\beta}^i$ the linear estimator given by

$$\hat{\beta}^i = \frac{1}{\lambda_i} v_i v_i' X'V^{-1}Y . \qquad (2.0.1)$$

We then have:

$$\hat{\beta} = \sum_{i=1}^{p} \hat{\beta}^i . \qquad (2.0.2)$$

And:

$$E_\beta[\hat{\beta}^i] = \frac{1}{\lambda_i} v_i v_i' X' V^{-1} X(\sum_{i=1}^{p} b_i v_i) = b_i v_i . \qquad (2.0.3)$$

The estimators $\hat{\beta}^i$ are called principal component estimators because they satisfy (2.0.3). One of the main reasons to use the GLSE in estimation problems is the fact that it is the unique linear estimator which satisfies:

$$E_\beta[\hat{\beta}] = \beta , \qquad \forall \beta \in \mathbb{R}^p$$

$$Var[\hat{\beta}] \leq Var[M], \quad \forall MY \in L \text{ such that } E_\beta[M] = \beta, \forall \beta \in \mathbb{R}^p .$$

If A and Q are symmetric matrices, $A \geq Q$ means that A-Q is a non-negative definite (n.n.d) matrix. Because $\hat{\beta}$ has those two properties it is often called the best linear unbiased estimator of β (BLUE of β).

Another subset of L which is important is the one formed by the restricted BLUE's, which are the BLUE's of β, when β is restricted to lie in some subspace S of $\mathbb{R}^p$, i.e. $\hat{\beta}^s$ is an S-restricted BLUE if:

$$E_\beta[\hat{\beta}^s] = \beta , \qquad \forall \beta \in S$$

$$Var[\hat{\beta}^s] \leq Var[M], \quad \forall MY \in L \text{ such that } E_\beta[M] = \beta, \forall \beta \in S.$$

The function $E[M] : \mathbb{R}^p \to \mathbb{R}^p$ defined as $E[M](\beta) = E_\beta[M]$ for all $\beta \in \mathbb{R}^p$ will be called expectation function of (the linear estimator) MY.

The function $B[M, \ ] : \mathbb{R}^p \to \mathbb{R}^p$ defined as $B[M,\beta] = \beta - E[M](\beta)$ for all $\beta \in \mathbb{R}^p$ will be called the <u>bias function of MY</u>. With statements of the kind "MY is a biased estimator of β if $E_\beta[M] \neq \beta$ for some $\beta \in \mathbb{R}^p$" or "MY is an unbiased estimator of β if $E_\beta[M] = \beta$ for all $\beta \in \mathbb{R}^p$" the explicit use of the expectation and bias functions is avoided. Nevertheless in contrast to this tradition we will make extensive use of both in the thesis. Some reasons for this choice are: When $MY \in L$, the expectation function of MY has a particularly simple form: $E[M] = MX$. We feel that those concepts give more freedom to state some propositions like the generalization of the Gauss-Markov Theorem given in the first section. A third reason is that when we want to give a description of the bias of an S-restricted BLUE on $\mathbb{R}^p$, which goes beyond the trivial "$E_\beta[\hat{\beta}^S] \neq \beta$ for some $\beta \in \mathbb{R}^p$" we are somehow compelled to speak of the expectation function of the S-restricted BLUE. This last point has been one of the motivations for this chapter and Corollary 2-1 gives a geometrical description of $B[M,\beta]$.

Another motivation for the results of the first section is the following one: as it was pointed out above, the BLUE can be expressed as the sum of particular estimators called principal component estimators. This representation has its advantages and has been exploited (as it will be seen later) in the context of generalized ridge estimators. The fact is that an analogous representation for S-restricted BLUE's is not available and in general the representation of an S-restricted BLUE is relatively complicated. A unified representation is therefore useful. As a result of this one is led to introduce the concept of <u>component estimator</u> which seems to be the natural generalization of the principal component estimator. This concept, besides having a clear statistical interpretation will prove to be very useful from a mathematical point of view for further developments in the thesis. It is somehow appealing that both motivations lead to the same kind of results.

Since 1970, arising out of the works of Hoerl and Kennard, other kinds of linear estimators have been used for estimation problems. They introduced the ridge estimators which are estimators of the form:

$$M_k Y = (X'V^{-1}X + kI)^{-1}X'V^{-1}Y , \qquad (2.0.4)$$

where k is a positive number. Those estimators are advantageous from at least two points of view:

The first one is that the inversion of $X'V^{-1}X + kI$ if k is "large enough" is a lot easier than the inversion of $X'V^{-1}X$ when $X'V^{-1}X$ is an ill conditioned matrix.

The second one, which is more relevant from a statistical point of view, is that for any $k > 0$ there exists a set $C_k \subset \mathbb{R}^p$ such that:

$$E_\beta[(M_k-\beta)'(M_k-\beta)] < E_\beta[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)] , \qquad \forall\, \beta \in C_k . \qquad (2.0.5)$$

In current literature $E_\beta[(M-\beta)'(M-\beta)]$ is denoted by $MSE[M,\beta]$ and is the mean square error of MY at β. (2.0.5) can be written as:

$$MSE[M,\beta] < MSE[\hat{\beta},\beta] , \qquad \forall\, \beta \in C_k .$$

An estimator which satisfies (2.0.5) is said to be better than the GLSE $\hat{\beta}$ on C_k (in terms of the mean square error).

The second property of the ridge estimators aroused an interest in the MSE properties of the elements of L. In this context the notion of minimum mean square error linear estimator (MMSELE) or best linear estimator (BLE) appeared (see Theil 1971, Rao 1971, Bibby 1972, Farebrother 1975). $MY \in L$ is a MMSELE (or BLE) if there exist $\beta_M \in \mathbb{R}^p$ such that

$$MSE[M,\beta_M] \leq MSE[N,\beta_M] \quad , \qquad \forall \; NY \in L \quad .$$

The second section of the chapter studies some properties of this kind of functions and it is shown that they are closely related to the component estimators discussed in the previous section.

Another property, which in some sense is in contrast to the above one, but closely related, is the notion of admissibility when the set of estimators considered for the definition of admissibility is L. Since this notion will play an essential part in the thesis, a formal definition of it is made:

DEFINITION 2.1.

Let $MY \in L$. MY is an admissible linear estimator for the model $(X,V,\mathbb{R}^p)$ (i.e.: MY is an ALE) if and only if there does not exist $NY \in L$ and $\beta_0 \in \mathbb{R}^p$ such that

$$MSE[N,\beta] \leq MSE[M,\beta] \; , \qquad \forall \; \beta \in \mathbb{R}^p \; .$$

and

$$MSE[N,\beta_0] < MSE[M,\beta_0] \qquad .$$

Shinozaki in 1975 studied this property and Rao (1976) gave the following characterization: $MY \in L$ is an ALE if and only if

$$M = AX'V^{-1}, \;\; A \text{ is symmetric and } AX'V^{-1}XA \leq A. \qquad (2.0.6)$$

(Here again $\leq$ is the partial order in the set of non negative symmetric matrices). Although this characterization is very clear from a mathematical point of view it has at least one important disadvantage from a statistical point of view. The bias of MY is given by:

$$B[M,\beta] = (I - AX'V^{-1}X)\,\beta \,. \qquad (2.0.7)$$

From this expression and (2.0.6) it is not easy to have a geometrical description of the bias. This is perhaps one of the main reasons to prefer using in practice the ridge estimator and some of its generalizations. In fact using the spectral decomposition of $X'V^{-1}X$, M_k can be written as:

$$M_k = \sum_{i=1}^{p} \frac{1}{\lambda_i+k} v_i v_i' X'V^{-1} \quad ,$$

or

$$M_k Y = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i+k} \hat{\beta}^i \quad ,$$

where the $\hat{\beta}^i$ are the principal component estimators and λ_i are the corresponding eigenvalues of $X'V^{-1}X$. If $\beta = \sum_{i=1}^{p} b_i v_i$ we have that the bias function of $M_k Y$ is:

$$B[M_k,\beta] = \sum_{i=1}^{p} \left(1 - \frac{\lambda_i}{\lambda_i+k}\right) b_i v_i \quad ,$$

which has a clearer geometrical interpretation than $B[M,\beta]$ in (2.0.7). The "spectral form" of the ridge estimator suggested these generalizations:

$$M_{\underset{\sim}{k}} Y = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i+k_i} \hat{\beta}^i \quad , \qquad k_i \geq 0, \qquad 1 \leq i \leq p,$$

and

$$M_{\underset{\sim}{\delta}} Y = \sum_{i=1}^{p} \delta_i \hat{\beta}^i \quad , \qquad 0 \leq \delta_i \leq 1,\ 1 \leq i \leq p.$$

Those estimators have been studied by Mayer and Willke 1973, Goldstein and Smith 1974, Rolph 1976 and Obenchain 1978 among others. They are usually called: "shrunken estimators", "shrinkage estimators", or "generalized ridge estimators". The δ_i are called shrinkage factors. One common feature to those estimators is that all of them are ALE and their bias function has the general form

$$B[M_{\underset{\sim}{\delta}},\beta] = \sum_{i=1}^{p} (1 - \delta_i)\, b_i v_i \quad ,$$

which again is easily interpretable in geometrical terms.

When in (2.0.4), kI is replaced by a non negative definite symmetric matrix G an important generalization of the ridge estimators is obtained; the estimators of the form:

$$M_G Y = (X'V^{-1}X + G)^{-1}X'V^{-1}Y.$$

For each $\underset{\sim}{k} = (k_1,\ldots,k_p)$, $k_i \geq 0$, $0 \leq i \leq p$, there exists an n.n.d.s matrix G such that

$$M_{\underset{\sim}{k}} Y = M_G Y \quad .$$

The converse is not true. Any estimator $M_G Y$ is an ALE and for any ALE MY there exist a sequence $\{G_n\}$ of n.n.d.s. matrices such that,

$$M_{G_n} \to M , \qquad (\text{as } n \to \infty) \quad .$$

Despite their explicit form (as opposed to the Rao's characterization given in (2.0.6)) the estimators $M_G Y$ lack in general an interpretable form of their bias function:

$$B[M_G,\beta] = (I - (X'V^{-1}X + G)^{-1}X'V^{-1}X)\ \beta.$$

In the third section of the chapter it is shown that any ALE MY can be represented as the sum of appropriate MMSELE. This representation is analogous to the one given in the first section for the restricted BLUE's in terms of component estimators. It allows us to give an expression for the bias function which has the same geometrical interpretation as the one for shrinkage estimators.

2.1 THE SET OF S-BLUE's

It is common in practice to be working with a linear model $(X,V,\mathbb{R}^p)$ with linear restrictions of the form $H\beta = 0$. When those restrictions apply we are working with a linear model (X,V,S) where $S = \{\beta | H\beta = 0\}$ and in this case the GLSE $\hat{\beta}$ corresponding to the "full model" $(X,V,\mathbb{R}^p)$ is not any more the BLUE and the restricted GLSE $\hat{\beta}^S$ corresponding to the "restricted model" (X,V,S) is better than $\hat{\beta}$ in the sense that

$$\mathrm{Var}[\hat{\beta}^S] \leq \mathrm{Var}[\hat{\beta}] , \qquad (\text{if } S \neq \mathbb{R}^p) .$$

We will now introduce some notation that will soon prove to be useful. Given a function f and a subset C of its domain, the restriction of f to C will be written f_C. In general when a function appears in the text with a subscript which is being used to name a set it will be understood that the domain of the function is the set denoted by the subscript; when no set subscript appears, the domain will be assumed to be $\mathbb{R}^p$ unless the contrary is specified. When two functions f and g coincide on a set C, we will write $f_C = g_C$. Let $f = E[M]$, then $E[M]_C = f_C$; i.e.: the restriction to C of the expectation function of MY will be denoted by $E[M]_C$. The letter S will be used to name sets which are subspaces. The next definition will play an essential part in the thesis, it is a convenient generalization of the definition of BLUE.

DEFINITION 2.2.

Let S a subspace of $\mathbb{R}^p$ and f a function, $f : \mathbb{R}^p \to \mathbb{R}^p$. Let $MY \in L$. Then MY is an <u>S-BLUE of f</u> if and only if:

$$E[M]_S = f_S ,$$

and

$$\mathrm{Var}[M] \leq \mathrm{Var}[N], \qquad \forall\ NY \in L \text{ such that } E[N]_S = f_S.$$

DEFINITION 2.2'.

When in the above definition f is the identity on $\mathbb{R}^p$, i.e.: $f = I$, the S-BLUE of f will be called S-BLUE and will be denoted by $\hat{\underline{\beta}}^s$. The $\mathbb{R}^p$- BLUE will be written $\hat{\beta}$.

The next theorem has been used by Seely and Zyskind (1971) to build up their theory of minimum variance linear unbiassed estimators and it is due to Lehmann and Scheffé (1950) (Theorem 5.3).

LEHMANN-SCHEFFÉ THEOREM.

Let (X,V,S) a linear model. Let $L_0^S = \{NY \in L \mid E[N]_S = 0\}$. Then MY is an S-BLUE for $f_S = E[M]_S$ if and only if $\mathrm{Cov}[M,N] = 0$ for all $NY \in L_0^S$.

PROOF. If MY is an S-BLUE of f_S we have that $\mathrm{Var}[M+N] \geq \mathrm{Var}[M]$ for all $NY \in L_0^S$ or $2\,\mathrm{Cov}[M,N] + \mathrm{Var}[N] \geq 0$ for all $NY \in L_0^S$; this implies in particular that $2t\,\mathrm{Cov}[M,N] + t^2\,\mathrm{Var}[N] \geq 0$ for all $t \in \mathbb{R}$, but this is only possible if $\mathrm{Cov}[M,N] = 0$. Conversely suppose that $\mathrm{Cov}[M,N] = 0$, then $\mathrm{Var}[M+N] = \mathrm{Var}[M] + \mathrm{Var}[N] \geq \mathrm{Var}[M]$, but this implies that MY is an S-BLUE of f_S because any estimator $M_0Y \in L$ such that $E[M_0]_S = f$ must be of the form $M_0Y = MY + NY$ for some $NY \in L_0^S$.

The next propositions will prove the general Gauss-Markov Theorem which constitutes essentially Theorem 2.1.

PROPOSITION 2.1.

Let a linear model $(X,V,\mathbb{R}^p)$ and $LY \in L$. Let S a subspace of $\mathbb{R}^p$, $\dim S = r$. Then:

(i) There exists an S-BLUE $M_L Y$ of $E[L]_S$.

(ii) $M_L Y = AX'V^{-1}Y$ and the rows of A are in S.

PROOF. Let M be the space of $1\times\ell$ matrices with the inner product $<N_1,N_2> = N_1VN_2'$. For linear estimators N_1Y, N_2Y we have $Cov[N_1,N_2] = <N_1,N_2>$ and $Var[N_1] = <N_1,N_1>$. Let $M_0^S \subset M$ such that if $N \in M_0^S$ then $E[N]_S = 0$. Let $\{N_i\}_{i=1}^n$ an orthonormal basis of M_0^S, L_j a row of L let $(M_L)_j = L_j - \sum_{i=1}^n <L_j,N_i> N_i = L_j - P_0^V(L_j)$ (here P_0^V is the projection according to the inner product <,> of M onto M_0^S). From its definition it is easy to see that $<(M_L)_j,N> = 0$ for all $N \in M_0^S$ and so $Cov[M_L,M] = 0$ for all $MY \in L_0^S$. The L.S. Theorem tells us that $M_L Y$ is an S-BLUE of $E[M_L]_S$. To prove (ii), first notice that $NY \in L_0^S$ iff the rows of N are orthogonal to X(S). Let $MY \in L$, the L.S. Theorem tells us that MY is S-BLUE of the restriction to S of its own expectation function iff $MVN' = 0$ for all $NY \in L_0^S$. This is equivalent to asking that $m_i'Vz = 0$ for all $z \in X(S)^{\perp}$ and for all rows m_i' of M. This in turn is equivalent to $Vm_i \in X(S)$ for $i = 1,\ldots,p$ or $m_i' = \alpha_i'X'V^{-1}$ for some $\alpha_i \in S$, $i = 1,\ldots,p$. This proves (ii).

PROPOSITION 2.2.

Let S a subspace of $\mathbb{R}^p$. Let M_1Y, $M_2Y \in L$ such that:

(i) $E[M_1]_S = E[M_2]_S$.

(ii) $Var[M_1] = Var[M_2] \leq Var[M]$ for all $MY \in L$ such that

$E[M]_S = E[M_1]_S$.

Then: $M_1 = M_2$.

PROOF. From (i) we have that $E[M_1-M_2]_S = 0$. Since $Cov[M_1-M_2,N] = Cov[M_1,N] - Cov[M_2,N]$ and $Cov[M_1,N] = Cov[M_2,N] = 0$ for all $NY \in L_0^S$ (L.S. Theorem) we have that $Cov[M_1-M_2,N] = 0$ for all $NY \in L_0^S$ and then again from the L.S. Theorem we can conclude that

$(M_1-M_2)Y$ is an S-BLUE for $f_S = 0$. Since $MY = 0$ (for all $Y \in \mathbb{R}^\ell$) is an S-BLUE of $f_S = 0$ and $Var[M] = 0$ we must have that $Var[M_1-M_2] = 0$ which is equivalent to $(M_1-M_2)V(M_1-M_2)' = 0$; but this is only possible if $M_1-M_2 = 0$, because V is a p.d. matrix.

PROPOSITION 2.3.

For any subspace S of $\mathbb{R}^p$, there exists a unique S-BLUE.

PROOF. The uniqueness is assured by the previous proposition. The existence from the fact that $E[\hat{\beta}]_S = I_S$. The next proposition is important to prove Theorem 2.1 and Theorem 2.2.

PROPOSITION 2.4.

Let $MY = AX'V^{-1}Y$ and the rows of A span a subspace S. Then MY is the S-BLUE of $f_S = E[M]_S$.

PROOF. From Prop. 2.1, the S-BLUE of $E[M]_S$ has the form $M_1Y = A_1X'V^{-1}Y$ with the rows of A_1 in S. Let $T = X'V^{-1}X$, we then have $E_\beta[M] = AT\beta = A_1T\beta$ for all $\beta \in S$. This implies that $(A-A_1)T\beta = 0$ for all $\beta \in S$ and the rows of $A-A_1$ are in S. Since T is a p.d.s. matrix this is only possible if $A-A_1 = 0$; or $A = A_1$.

The next theorem, which summarizes the results contained in the preceding propositions, is a generalization of the Gauss-Markov Theorem which we have not found in the literature.

THEOREM 2.1

Let a linear model $(X,V,\mathbb{R}^p)$. Let $MY \in L$ such that $MY = AX'V^{-1}Y$. Let S a subspace of $\mathbb{R}^p$. Then: MY is the unique S-BLUE of $E[M]_S$ if and only if the rows of A are in S.

To start the study of S-BLUE's we will consider the simplest case, the case when dim S = 1.

DEFINITION 2.3.

MY is called a <u>component estimator</u> if there exists a subspace S of dimension 1 such that MY = $\hat{\beta}^S$ (i.e. MY is the S-BLUE, see Definition 2.2').

Given a p.d.s. matrix T we will say that β is T-orthogonal to γ iff $\beta' T\gamma = 0$. We will write $\beta \perp_T \gamma$. The set $\{\beta_i\}_{i=1}^n$ is a set of T-orthonormal vectors iff $\beta_i' \, T \, \beta_j = \delta_{ij}$ (here δ_{ij} denotes the Kronecker delta). $B = \{\beta_i\}_{i=1}^n$ is a T-orthonormal basis of S iff B is a basis of S and B is a set of T-orthonormal vectors. Also the T-norm of β will be $||\beta||_T = (\beta' T\beta)^{1/2}$. The T-orthogonal complement of S is the set of T-orthogonal vectors to S and will be denoted by $S^{\perp T}$. When T = I, the T will be dropped in the above definitions and notation. The following proposition gives an explicit expression for component estimators.

PROPOSITION 2.5.

Let $\beta \in \mathbb{R}^p$ and $S = \text{Span}\,\{\beta\}$. Let $T = X'V^{-1}X$. Then:

$$\hat{\beta}^S = \frac{1}{||\beta||_T^2} \beta\beta' X' V^{-1} Y \quad .$$

PROOF. Let $\gamma \in S$ and $\lambda \in \mathbb{R}$ such that $\gamma = \lambda\beta$. By Prop. 2.1 $\hat{\beta}^S = \alpha\beta' X' V^{-1} Y$ for some vector $\alpha \in \mathbb{R}^p$, because any matrix with rows in S has the form $\alpha\beta'$ if $S = \text{Span}\,\{\beta\}$. From this we must have that $E_\gamma[\hat{\beta}^S] = \alpha\beta' X' V^{-1} X\gamma = \lambda ||\beta||_T^2 \alpha$. From the definition of S-BLUE we must also have that $E_\gamma[\hat{\beta}^S] = \gamma = \lambda\beta$. This implies that $\alpha = \frac{1}{||\beta||_T^2}\beta$ and the proposition is proved.

If v_i is an eigenvector of T, we have that

$$Tv_i - \lambda_i v_i = 0 ,$$

or

$$v_i'Tv_i - \lambda_i v_i'v_i = 0 ,$$

and then

$$||v_i||_T^2 = \lambda_i ||v_i||^2 .$$

When v_i is normalized

$$||v_i||_T^2 = \lambda_i .$$

Then if in Prop 2.5, β is replaced by v_i, the principal component estimator $\hat{\beta}^i$ is obtained. This justifies partially the name of component estimator for $\hat{\beta}^S$ when dim S = 1.

Let $P_S^T : \mathbb{R}^p \to \mathbb{R}^p$ a linear function such that $P_S^T(\beta) = \beta$ if $\beta \in S$ and $P_S^T(\beta) = 0$ if $\beta \in S^{\perp T}$. P_S^T is in fact the projection of $\mathbb{R}^p$ onto S when the inner product used is given by $\langle\beta,\gamma\rangle = \beta'T\gamma$. We will call it T-projection (of $\mathbb{R}^p$ onto S). From now on $T = X'V^{-1}X$ unless otherwise specified. A simple calculation shows that if $\{\beta_i\}_{i=1}^r$ is a T-orthonormal basis of S (see after Definition 2.3 for the definition); we have that $P_S^T = M_S T = \sum_{i=1}^r \beta_i\beta_i'T$. Using this notation in Proposition 2.5 we see that $\hat{\beta}^S = M_S X'V^{-1}Y$ and it becomes evident that $E[\hat{\beta}^S] = P_S^T$. The next theorem shows that those properties are true for $\hat{\beta}^S$ even if dim S $\neq$ 1. The proof is immediate: let $MY = \sum_{i=1}^r \beta_i\beta_i'X'V^{-1}Y$, from Prop. 2.4, MY is the S-BLUE of $f_S = E[M]_S$ (because the rows of $\sum_{i=1}^r \beta_i\beta_i'$are inS); also $E_\beta[M] = \sum_{i=1}^r \beta_i\beta_i'T\beta = P_S^T(\beta)$ and then $E[M]_S = I_S$. Definition 2.2' implies then that MY is the S-BLUE.

THEOREM 2.2.

Let $(X,V,\mathbb{R}^p)$ a linear model. Let S a subspace of $\mathbb{R}^p$. Let $T = X'V^{-1}X$ and $\{\beta_i\}_{i=1}^r$ a T-orthonormal basis of S. Then

(i) $$\hat{\beta}^S = \sum_{i=1}^{r} \beta_i\beta_i'X'V^{-1}Y.$$

(ii) $$E[\hat{\beta}^S] = P_S^T .$$

As an immediate corollary we obtain a geometric interpretation for the bias of $\hat{\beta}^S$.

COROLLARY 2.2.1

The bias function of the S-BLUE is given by $B[\hat{\beta}^S] = I - P_S^T$.
Theorem 2.2, (i) suggests an "additive law" for S-BLUE's.

COROLLARY 2.2.2

Let S, S_1, S_2 subspaces of $\mathbb{R}^p$. Then $\hat{\beta}^S = \hat{\beta}^{S_1} + \hat{\beta}^{S_2}$ iff $S = S_1 \oplus_T S_2$ and $S_1 \perp_T S_2$.

PROOF. Let $B_1 = \{\beta_i\}_{i=1}^r$, $B_2 = \{\gamma_i\}_{i=1}^n$ T-orthonormal bases of S_1 and S_2 respectively, then $B_1 \cup B_2$ is a T-orthonormal basis of S (if $S = S_1 \cup S_2$), but this implies as a consequence of Theorem 2.2, (i) that $\hat{\beta}^S = \hat{\beta}^{S_1} + \hat{\beta}^{S_2}$. The other implication can be seen as follows. Suppose that $\hat{\beta}^S = \hat{\beta}^{S_1} + \hat{\beta}^{S_2}$, then we have that $P_S^T(\beta) = E_\beta[\hat{\beta}^S] =$
$= E_\beta[\hat{\beta}^{S_1}] + E_\beta[\hat{\beta}^{S_2}] = P_{S_1}^T(\beta) + P_{S_2}^T(\beta)$ (from Theorem 2.2(ii)); but since P_S^T, $P_{S_1}^T$ and $P_{S_2}^T$ are projections this can only be possible if $S_1 \cup S_2 = S$ and $S_1 \perp_T S_2$.

COROLLARY 2.2.3.

Let $\{\beta_i\}_{i=1}^r$ a T-orthonormal basis of S and $\hat{\beta}^i$ the component estimator corresponding to β_i. Then:

$$\hat{\beta}^S = \sum_{i=1}^{r} \hat{\beta}^i \quad .$$

PROOF. It is immediate from Prop 2.5 and Theorem 2.2, (i). This corollary can be seen essentially as a generalization of the spectral decomposition of the GLSE. This further justifies the name of component estimators to the S-BLUE's of subspaces of dimension one.

The following remark perhaps helps to avoid possible confusion. Given an arbitrary basis $\{\gamma_i\}_{i=1}^r$ of S, which is not T-orthonormal or even T-orthogonal, the S-BLUE can always be decomposed as the addition of estimators $\hat{\gamma}^i$ (i.e. $\hat{\beta}^S = \sum_{i=1}^{r} \hat{\gamma}^i$) such that if $\beta = \sum_{i=1}^{r} b_i\gamma_i$ we have $E_\beta[\hat{\gamma}^i] = b_i\gamma_i$. The existence of the $\hat{\gamma}^i$ can be seen in the following way; for $\{\gamma_i\}_{i=1}^r$ there exist always a p.d. matrix π such that the π-projection P_i^π of $\mathbb{R}^p$ onto $S_i = \text{Span}\{\gamma_i\}$ satisfies $P_i^\pi(\beta) = b_i$. We then have that $\beta = \sum_{i=1}^{r} P_i^\pi(\beta)\gamma_i$ and $\hat{\gamma}^i = P_i^\pi \hat{\beta}^S$. Now only if we further require that $P_i^\pi\hat{\beta}^S$ is the S_i-BLUE ($1 \le i \le r$) the set $\{\gamma_i\}_{i=1}^r$ must be T-orthonormal.

COROLLARY 2.2.4.

The variance of an S-BLUE $\hat{\beta}^S = \sum_{i=1}^{r} \beta_i\beta_i'X'V^{-1}Y$ is

$$\text{Var}[\hat{\beta}^S] = \sum_{i=1}^{r} \beta_i\beta_i' \quad .$$

2.2. THE SET OF MINIMUM MEAN SQUARE ERROR LINEAR ESTIMATORS

Minimum mean square error linear estimators (MMSELE) have been studied, among others, by Theil (1971), Rao (1971), Bibby (1972), Farebrother (1975). In this section a new way of deriving MMSELE's will be given. Also some properties of MMSELE's are obtained. In our context the most important property of MMSELE's is their connection with the component estimators which is summarized in Proposition 2.6. At the end of the section a subset of L is defined by analogy with the set of S-BLUE's.

DEFINITION 2.4.

Let $(X,V,\mathbb{R}^p)$ a linear model. Let $\beta \in \mathbb{R}^p$. The minimum mean square error linear estimator (MMSELE) of β, $M_\beta Y$ is the linear estimator which satisfies:

$$MSE[M_\beta, \beta] \leq MSE[M,\beta] \quad , \qquad \forall\ MY \in L . \tag{2.2.1}$$

To be able to state some properties of this section and of the next chapters we will introduce more general notions of "quadratic risks".

DEFINITION 2.5.

Let Q a n.n.d. matrix. The Q quadratic risk of MY at β will be

$$R_Q[M,\beta] = E_\beta[(M-\beta)'Q(M-\beta)].$$

It is immediate that:

$$R_I[M,\beta] = MSE[M,\beta] \quad .$$

The Q-bias of MY at β will be:

$$B_Q[M,\beta] = ||E_\beta[M] - \beta||_Q^2.$$

The Q-variance of MY at β will be:

$$\mathrm{Var}_Q[M] = \mathrm{Tr}[Q\,\mathrm{Var}[M]] \quad .$$

(Here Tr[A] denotes Trace of A.)

We shall often use the well known relationship between R_Q, B_Q and Var_Q:

$$R_Q[M,\beta] = B_Q[M,\beta] + \mathrm{Var}_Q[M]. \qquad (2.2.2)$$

Now let S = Span{β} and P_S^Q the Q-projection of $\mathbb{R}^p$ onto S (we consider that Q is a p.d.s. matrix). We then have:

$$B_Q[M,\beta] = ||E_\beta[M] - \beta||_Q^{\,2} = ||(MX-I)\beta||_Q^{\,2}$$

$$\geq ||P_S^Q (MX-I)\beta||_Q^{\,2} = ||E_\beta[P_S^Q M] - \beta||_Q^{\,2} = B_Q[P_S^Q M,\beta] , \qquad (2.2.3)$$

and if $H = S^{\perp Q}$:

$$\mathrm{Var}_Q[M] = \mathrm{Tr}[QMVM'] = \mathrm{Tr}[Q(P_S^Q + P_H^Q)MVM']$$

$$= \mathrm{Tr}[QP_S^Q MVM'] + \mathrm{Tr}[QP_H^Q MVM']$$

$$= \mathrm{Tr}[QP_S^Q MVM'(P_S^Q)'] + \mathrm{Tr}[QP_H^Q MVM'(P_H^Q)']$$

$$\geq \mathrm{Tr}[QP_S^Q MVM'(P_S^Q)'] = \mathrm{Var}_Q[P_S^Q M] \quad ; \qquad (2.2.4)$$

(2.2.3) and (2.2.4) imply that if $M_\beta Y$ satisfies:

$$R_Q[M_\beta,\beta] \leq R_Q[M,\beta] , \qquad \forall\, MY \in L \; .$$

$M_\beta Y$ must also satisfy $E_\beta[M_\beta] = \lambda\beta$ for some $\lambda \in \mathbb{R}$. Since $\mathrm{Var}_Q[\lambda\hat{\beta}^S] \leq \mathrm{Var}_Q[M]$ for all $MY \in L$ such that $E_\beta[M] = \lambda\beta$ we have that:

$$M_\beta Y = \lambda\hat{\beta}^S \; , \qquad \text{for some } \lambda \in \mathbb{R} \; .$$

Also:

$$R_Q[M_\beta,\beta] = ||(\lambda\frac{\beta\beta'}{||\beta||_T^2} T-I)\beta||_Q^2 + Tr[Q(\lambda\frac{\beta\beta'}{||\beta||_T^2} \quad T \frac{\beta\beta'}{||\beta||_T^2} \lambda)] \quad ,$$

$$R_Q[M_\beta,\beta] = (\lambda-1)^2 \; ||\beta||_Q^2 + \lambda^2 \frac{||\beta||_Q^2}{||\beta||_T^2} \quad .$$

Then if $M_\beta Y$ satisfies (2.2.1), we must have:

$$\frac{d}{d\lambda} (\lambda-1)^2||\beta||^2 + \lambda^2 \frac{||\beta||^2}{||\beta||_T^2} = 0 \quad .$$

Or

$$\lambda - 1 + \frac{\lambda}{||\beta||_T^2} = 0 \quad .$$

From this we obtain

$$\lambda = \frac{||\beta||_T^2}{1 + ||\beta||_T^2} \quad .$$

Since this result is independent of I, it is also true that for any p.d.s matrix Q we have:

$$R_Q[M_\beta,\beta] \le R_Q[M,\beta] \quad , \qquad \forall \; MY \in L \; .$$

This result was quoted by Bibby 1972. The proof we have given is shorter and avoids matrix differentiation. The result can be extended to n.n.d.s matrices Q in the following way: let Q_1 such that $Q_\varepsilon = Q + \varepsilon Q_1$ is a p.d.s. matrix for all $\varepsilon > 0$. The result follows from the fact that:

$$R_Q[M,\beta] = \lim_{\varepsilon\to 0} R_{Q_\varepsilon}[M,\beta], \qquad \forall \; MY \in L \; .$$

We obtain in particular that

$$R_{\alpha\alpha'}[M_\beta,\beta] \leq R_{\alpha\alpha'}[M,\beta] \qquad \forall\, \alpha \in \mathbb{R}^p,\ \forall\, MY \in L \ .$$

Since

$$\alpha' E_\beta[(M_\beta-\beta)(M_\beta-\beta)']\alpha = R_{\alpha\alpha'}[M,\beta] \ .$$

We have that

$$E_\beta[(M_\beta-\beta)(M_\beta-\beta)'] \leq E_\beta[(M-\beta)(M-\beta)'] \ , \quad \forall\, MY \in L.$$

We also have

$$B_Q[M_\beta,\beta] = (1-\lambda)^2 ||\beta||_Q^{\ 2} = \frac{||\beta||_Q^{\ 2}}{(1+||\beta||_T^{\ 2})^2} \ ,$$

$$\mathrm{Var}_Q[M_\beta,\beta] = \lambda^2 \frac{||\beta||_Q^2}{||\beta||_T^2} = \frac{||\beta||_T^2 ||\beta||_Q^2}{(1+||\beta||_T^2)^2} \ ,$$

and

$$R_Q[M_\beta,\beta] = \frac{||\beta||_Q^2}{1+||\beta||_T^2} \ .$$

To be able to refer to those results we will put them in form of propositions.

PROPOSITION 2.6.

(i) $MY \in L$ is a MMSELE (for some unknown β) iff it has the form $MY = \delta\hat{\beta}^S$, $0 \leq \delta < 1$ and $\hat{\beta}^S$ is a component estimator.

(ii) The MMSELE for β, $M_\beta Y$ is given by

$$M_\beta Y = \frac{||\beta||_T^{\ 2}}{1+||\beta||_T^{\ 2}} \ \hat{\beta}^S \ .$$

Where $S = \mathrm{Span}\ \{\beta\}$.

PROPOSITION 2.7.

(i) Let $\alpha \in \mathbb{R}^p$. Then:

$$R_{\alpha\alpha'}[M_\beta,\beta] \leq R_{\alpha\alpha'}[M,\beta], \qquad \forall\, MY \in L .$$

(ii) Let Q a n.n.d.s. matrix then

$$R_Q[M_\beta,\beta] \leq R_Q[M,\beta], \qquad \forall\, MY \in L.$$

(iii) $E_\beta[(M_\beta-\beta)(M_\beta-\beta)'] \leq E_\beta[(M-\beta)(M-\beta)'], \quad \forall\, MY \in L$.

PROPOSITION 2.8.

For any $\beta \in \mathbb{R}^p$, and Q n.n.d.s. matrix we have:

(i) $$B_Q[M_\beta,\beta] = \frac{||\beta||_Q^2}{(1+||\beta||_T^2)^2} .$$

(ii) $$\mathrm{Var}_Q[M_\beta] = \frac{||\beta||_T^2||\beta||_Q^2}{(1+||\beta||_T^2)^2} .$$

(iii) $$R_Q[M_\beta,\beta] = \frac{||\beta||_Q^2}{1+||\beta||_T^2} .$$

The MMSELE have no practical value mainly because $B_Q[M_\beta,\beta_0]$ grows fast as soon as β_0 is "not near" to β. They should be considered mainly as theoretical objects. By analogy to the shrinkage estimators and in the light of Theorem 2.2 and Proposition 2.6 it is natural to ask about the properties of estimators of the form

$$MY = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y ,$$

where $0 \leq \delta_i \leq 1$, $i = 1,\ldots,p$ and $\{\beta_i\}_{i=1}^p$ is a T-orthonormal basis of $\mathbb{R}^p$. It will be proved in the next section that these estimators form in fact the set of admissible linear estimators of the linear model $(X,V,\mathbb{R}^p)$.

2.3. THE SET OF ADMISSIBLE LINEAR ESTIMATORS

The next Lemma will play an essential role in most of the results of the thesis. The statements given in it are given in enough generality to cope with the different applications in which it will be used.

LEMMA 2.1.

Let T be an arbitrary n.n.d.s matrix (not necessarily $X'V^{-1}X$); A a symmetric matrix, such that $N(A)^{\perp} \subset N(T)^{\perp} = S$.

Let A_S, A_H and A_J the restriction of A to S, H and J respectively; where $H = S \cap (S \cap N(A))^{\perp T_S}$, $J = \text{Range } A = N(A)^{\perp}$ and $h = \dim A(S) = \dim H = \dim J$. Then there exist a basis $B_1 = \{\beta_i\}_{i=1}^{h}$ of H and a basis $B_2 = \{\gamma_i\}_{i=1}^{h}$ of J, such that

(a) $\beta_i' T \beta_j = \gamma_i' T \gamma_j = \delta_{ij}$ $\quad i,j = 1,\ldots,h.$

(b) $A = T(\sum_{i=1}^{h} \lambda_i \beta_i \beta_i')T$ $\quad \lambda_i = \beta_i' A \beta_i$.

(c) $A = \sum_{i=1}^{h} \mu_i \gamma_i \gamma_i'$, $\mu_i = (\gamma_i' A_J^{-1} \gamma_i)^{-1}$ and A_J^{-1} is the inverse of the restriction of A to J.

(d) $A_H^{-1} = (\sum_{i=1}^{h} \frac{1}{\lambda_i} \beta_i \beta_i')_J$.

(e) $A_J^{-1} = (T(\sum_{i=1}^{h} \frac{1}{\mu_i} \gamma_i \gamma_i')T)_J$.

(f) Let T_S^{-1} be the inverse of the restriction of T to S. Let $r = \dim S$ and $\{\varepsilon_i\}_{i=1}^{r}$ any T_S-orthonormal basis of S. Then $T_S^{-1} = (\sum_{i=1}^{r} \varepsilon_i \varepsilon_i')_S$.

PROOF. Let $C_i = S \cap \{\beta \mid \|\beta\|_T^2 = 1\} \cap \{\beta_1, \ldots, \beta_{i-1}\}^{\perp T_S}$, then the vector β_i which satisfies

$$\max_{\beta \in C_i} \|\beta\|_A^2 = \|\beta_i\|_A^2 , \qquad (2.3.1)$$

can be found using Lagrange multipliers through the equation:

$$\frac{\partial}{\partial \beta} \beta' A \beta - \lambda_i(\beta' T \beta - 1) - \sum_{K=1}^{i-1} \nu_K (\beta' T \beta_K) - \sum_{K=1}^{\ell} \rho_K \beta' v_K = 0 , \qquad (2.3.2)$$

$\{v_K\}_{K=1}^{\ell}$ is a basis of $N(T)$.
(the existence of the solution comes from the fact that C_i is a compact set and $\|\beta\|_A^2$ is a continuous function of β). We will prove that $\beta_i \in C_i$ satisfies (2.3.1) if and only if the following equality is satisfied

$$(A - \lambda_i T)\beta_i = 0 . \qquad (2.3.3)$$

For $i = 1$ it is immediate that (2.3.3) reduces to (2.3.2). Suppose that the equivalence is true for $i-1$. It is obvious that (2.3.3) and $\beta_i \in C_i$ implies (2.3.2). To see the other implication notice first that (2.3.2) implies that $(A - \lambda_i T)\beta_i$ belongs to $S_i = \text{Span}\{T\beta_k\}_{k=1}^{i-1}$ because $\beta_i \in S$ and then $(A - \lambda_i T)\beta_i = \sum_{k=1}^{i-1} \nu_k T\beta_k$. If $(A - \lambda_i)T\beta_i \neq 0$ there is an $n < i$ such that $\nu_n \neq 0$ and $\beta_n' (A - \lambda_i T)\beta_i = \nu_n \sum_{k=1}^{i-1} \beta_n' T\beta_k = \nu_n$ (from the induction hypothesis); since $\beta_i \in C_i$ we must have $\beta_n' T\beta_i = 0$ and then $\beta_n' A \beta_i \neq 0$. From the induction hypothesis, since $n < i$, we have that $\beta_n' A = \lambda_n \beta_n' T$ and then $\beta_n' A \beta_i = \lambda_n \beta_n' T\beta_i = 0$ (because $\beta_i \in C_i$); we have reached a contradiction; this proves that $(A_H - \lambda_i T)\beta_i = 0$. Now because $\beta_i \in C_i$ we have that $\|\beta_i\|_T^2 = 1$ and from (2.3.1) $\lambda_i = \beta_i' A \beta_i > 0$ if $i \leq h$. The set $\{\beta_i\}_{i=1}^h$ obtained in this way is called a set of T-eigenvectors of A_H. From its construction it is immediate that it satisfies $\beta_i' T\beta_j = \delta_{ij}$, $i,j = 1,\ldots,h$ and that it is a T_H-orthonormal basis of H. Now since

$v \in S$ can be written as $v = \sum_{i=1}^{h} b_i\beta_i + u$ where $u \in N(A)$, we have

$$A(v) = A(\sum_{i=1}^{h} b_i\beta_i) = A(\sum_{i=1}^{h} b_i\beta_i) =$$

$$= \sum_{i=1}^{h} b_i A(\beta_i) = \sum_{i=1}^{h} b_i\lambda_i T\beta_i = T(\sum_{i=1}^{h} b_i\lambda_i\beta_i) =$$

$$= T(\sum_{i=1}^{h} b_i\lambda_i\beta_i(\beta_i'T\beta_i)) = T(\sum_{i=1}^{h} \lambda_i\beta_i\beta_i'(Tb_i\beta_i))=$$

$$= T(\sum_{i=1}^{h} \lambda_i\beta_i\beta_i')T(\sum_{i=1}^{h} b_i\beta_i) = T(\sum_{i=1}^{h} \lambda_i\beta_i\beta_i')Tv$$

This proves (b) and part of (a). To see (d) we only need to see that $A_H A_H^{-1} = I_{A(H)}$, $A_H^{-1} A_H = I_H$. Those conditions can be checked using (a) by expressing the elements of A(H) and H as linear combinations of $\{T\beta_i\}_{i=1}^{h}$ and $\{\beta_i\}_{i=1}^{h}$ respectively.

To see (c), from the proved part of (a) and (b) we know that there exists a set $\{\gamma_i\}_{i=1}^{h}$ such that $A_J^{-1} = (T\sum_{i=1}^{h} r_i\gamma_i\gamma_i'T)_J$, $r_i = \gamma_i' A_J^{-1}\gamma_i$. By (d) we obtain $A_J = (A_J^{-1})^{-1} = (\sum_{i=1}^{h} \frac{1}{r_i}\gamma_i\gamma_i')_J$. By similar arguments to those used in (b) it can be seen that $A(v) = (\sum_{i=1}^{h} \frac{1}{r_i}\gamma_i\gamma_i')v$ for all $v \in \mathbb{R}^p$. This proves (c) and (e). If we make $A = T$, (f) follows immediately from (d).

The next theorem shows that the set of estimators defined at the end of Section 2.2 is the set of ALE's. This theorem will play an essential role in the thesis.

THEOREM 2.3.

Let a linear model $(X,V,\mathbb{R}^p)$. Let $MY \in L$. Then MY is an ALE for the linear model $(X,V,\mathbb{R}^p)$ if and only if

$$MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X' V^{-1} Y ,$$

where $\{\beta_i\}_{i=1}^p$ is a T-orthonormal basis of $\mathbb{R}^p$ and $0 \le \delta_i \le 1$, $1 \le i \le p$.

PROOF. The first point is that if MY is an ALE it must be the BLUE of its own expectation function. To see this, let $f(\beta) = E_\beta[M]$ and M_1Y the BLUE of $f(\beta)$, then we have that for any $\beta \in \mathbb{R}^p$

$$R_I[M_1,\beta] = B_I[M_1,\beta] + Var_I[M_1] ,$$

$$= B_I[M,\beta] + Var_I[M_1] < B_I[M,\beta] + Var_I[M] = R_I[M,\beta] ,$$

and then MY is not an ALE. From Proposition 2.1 (ii) MY has the form $MY = AX'V^{-1}Y$. Now from Prop. 3.4 MY must be an ALE under the risk defined by T; but then Prop. 3.5 and 3.6 imply that $A = \sum_{i=1}^{p} \delta_i \beta_i \beta_i'$, where $\{\beta_i\}_{i=1}^p$ is a T-orthonormal basis of $\mathbb{R}^p$ and $0 \le \delta_i \le 1$ for $1 \le i \le p$. The other implication of the theorem is Prop. 4.8.

We will prove now as a corollary the characterization given by Rao, 1976.

COROLLARY 2.3.1.

MY is an ALE iff $MY = AX'V^{-1}Y$ and A is a symmetric matrix which satisfies $ATA \le A$.

PROOF. If MY is an ALE, from the theorem we know that $MY = AX'V^{-1}Y$, A is symmetric and $A = \sum_{i=1}^{p} \delta_i \beta_i \beta_i'$ where $\{\beta_i\}_{i=1}^p$ is a set of T-orthonormal vectors. Thus

$$(\sum_{i=1}^{p} \delta_i\beta_i\beta_i')T(\sum_{i=1}^{p} \delta_i\beta_i\beta_i') = \sum_{i=1}^{p} \delta_i^2\beta_i\beta_i' \leq \sum_{i=1}^{p} \delta_i\beta_i\beta_i'.$$

To obtain the other implication, from Lemma 2.1 (c) we know that there exists a set $\{\beta_i\}_{i=1}^{p}$ of T-orthonormal vectors such that $A = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'$. Let $\alpha_i' = \beta_i'T$, then $\alpha_i'ATA\alpha_i = \delta_i^2$ and $\alpha_i'A\alpha_i = \delta_i$, since $ATA \leq A$ we must have $\delta_i^2 \leq \delta_i$ but this implies that $0 \leq \delta_i \leq 1$. This completes the proof.

The next corollary provides us with a geometrical understanding of the bias function of an arbitrary ALE.

COROLLARY 2.3.2.

Let $MY = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y$ an ALE. Let $\beta = \sum_{i=1}^{p} b_i\beta_i$. Then the bias function of MY is

$$B[M,\beta] = \sum_{i=1}^{p} (1-\delta_i)b_i\beta_i .$$

We will use Lemma 2.1 to see how an estimator MY which is given in the form $MY = (T+G)^{-1}X'V^{-1}Y$ can be put in the form given in Theorem 2.3. Consider the ellipsoid $B_G = \{\beta | \beta'G\beta \leq 1\}$. By analogy to the case when $T = I$ the i^{th} T-main axis of B_G is defined to be just the set of vectors of the form $\pm\sqrt{\lambda}\beta_i$ with $\lambda \leq \frac{1}{\lambda_i}$ where β_i and λ_i are given by $G = T(\sum_{i=1}^{p} \lambda_i\beta_i\beta_i')T$ (in the case that G is singular, the T-main axis corresponding to $\lambda_i = 0$ will be the subspace spanned by β_i). Let γ_i be a vector of the subspace spanned by β_i, such that $\gamma_i'G\gamma_i = 1$; we have that $\| \gamma_i \|_T^2 = \frac{1}{\lambda_i}$ (if $\lambda_i \neq 0$) and

$$\frac{1}{1 + \lambda_i} = \frac{\| \gamma_i \|_T^2}{1 + \| \gamma_i \|_T^2} . \qquad (2.3.4)$$

From Lemma 2.1, (f) we have

$$TT^{-1}T = T\left(\sum_{i=1}^{p} \beta_i\beta_i'\right)T.$$

Therefore

$$T+G = T\left(\sum_{i=1}^{p} (1+\lambda_i)\,\beta_i\beta_i'\right)T.$$

From Lemma 2.1, (d)

$$(T+G)^{-1} = \sum_{i=1}^{p} \frac{1}{1+\lambda_i}\,\beta_i\beta_i' \quad .$$

From (2.3.4)

$$(T+G)^{-1} = \sum_{\lambda_i \neq 0} \frac{||\gamma_i||_T^{\,2}}{1+||\gamma_i||_T^{\,2}}\,\beta_i\beta_i' + \sum_{\lambda_i = 0} \beta_i\beta_i' \quad .$$

From Proposition 2.6

$$(T+G)^{-1}X'V^{-1}Y = \sum_{\lambda_i \neq 0} M\gamma_i Y + \sum_{\lambda_i = 0} \beta_i\beta_i'X'V^{-1}Y.$$

If G is a p.d. matrix all the λ_i are different from zero and

$$(T+G)^{-1}X'V^{-1}Y = \sum_{i=1}^{p} M\gamma_i\, Y.$$

i.e. $(T+G)^{-1}$ is the sum of the MMSELE corresponding to extreme points of the T-main axis of the ellipsoid B_G.

From the previous development it becomes clear that the only ALE which cannot be put in the form $(T+G)^{-1}X'V^{-1}Y$ for some n.n.d.s. matrix G are those which in the form $\sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y$ have $\delta_j = 0$ for some $1 \leq j \leq p$. The results are summarized in the next theorem.

THEOREM 2.4.

Let $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ an ALE.

(i) There exists a n.n.d.s. matrix G such that $MY = (T+G)^{-1}X'V^{-1}Y$ if and only if $\delta_i \neq 0$, $1 \leq i \leq p$. When such a G exists, it is unique and is given by

$$G = T\left(\sum_{i=1}^{p} \frac{1-\delta_i}{\delta_i} \beta_i \beta_i'\right) T.$$

(ii) Let G a p.d.s. matrix; let $\{\gamma_i\}_{i=1}^{p}$ a set of extreme points of the T-main axis of the ellipsoid B_G and $\{M\gamma_i Y\}_{i=1}^{p}$ the corresponding MMSELE. Then

$$(T+G)^{-1}X'V^{-1}Y = \sum_{i=1}^{p} M\gamma_i Y.$$

If G is not p.d. but only n.n.d, the equality remains valid if the M_{γ_i} corresponding to zero T-eigenvalues of G are replaced by the component estimators of their T-main axis.

From now on if $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ the set $\{\beta_i\}_{i=1}^{p}$ will be called a <u>set of axis</u> of MY and $\{\delta_i\}_{i=1}^{p}$ the set of <u>shrinkage factors</u> of MY. One set of axis and one set of shrinkage factors define a unique ALE. An ALE defines a unique set of shrinkage factors (including possible multiplicities) but in general it does not define a unique set of axis.

CHAPTER 3

In the previous chapter the inner product defined by the matrix $T = X'V^{-1}X$ has been extensively used. When X is not full rank the matrix T no longer defines an inner product on $\mathbb{R}^p$. Nevertheless it will be seen that the results obtained in Chapter 2 can be kept with just minor changes.

The chapter will have four sections:

(i) In the first section necessary and sufficient conditions are given for the existence of S-BLUE's. It is also proved that the representation for an S-BLUE in terms of component estimators (Theorem 2.2) remains valid when X is not full rank, if the S-BLUE exists.

(ii) The second section of the chapter deals with the connection of our formulation and the formulation in terms of g-inverses.

(iii) The third section characterizes the set of ALE when X is non-full rank. To prove this characterization some general propositions about risk inequalities are proved. Those results will also be useful in the next chapters.

(iv) The fourth section will give a characterization for admissible linear estimators when β is restricted to be in a subspace. In that section X will be assumed to be full rank.

3.1. THE EXISTENCE OF S-BLUE's

The first step in this section will be to extend Prop. 2.1 and Prop. 2.2 to the present situation.

PROPOSITION 3.1.

Let $LY \in L$. Let S a subspace of $\mathbb{R}^p$. Let $f_S = E[L]_S$. Then:

(i) There exists an S-BLUE of f_S;

(ii) Any S-BLUE of f_S has the form

$$MY = (A + M_1)X'V^{-1}Y,$$

with the rows of A in S and the rows of M_1 in N(X).

(iii) The S-BLUE of f_S is unique.

PROOF. (i) can be proved exactly the same way as Prop. 2.1(i). Following the line of the proof of (ii) of Prop 2.1 we can conclude that $Vm_i \in X(S)$ for all rows m_i' of M, but this is equivalent to asking that $m_i' = \alpha_i'X'V^{-1}$, $\alpha_i \in S$, $i = 1,\ldots,p$. Since V is invertible $m_i = V^{-1}X'\gamma_i$ if and only if $\gamma_i = \alpha_i + u_i$ where $u_i \in N(X)$. Then we obtain that $MY = (A + M_1)X'V^{-1}Y$ where the rows of A are given by the α_i' 's and the rows of M_1 by the u_i' 's. The proof of the unicity of the S-BLUE of f_S is the same as the one given in Prop. 2.2.

Without loss of generality we will always assume that $M_1 = 0$ because $MY = (A+M_1)X'V^{-1}Y = AX'V^{-1}Y$ for all $Y \in \mathbb{R}^\ell$. The next proposition is a generalization of Prop. 2.4.

PROPOSITION 3.2.

Let $MY = AX'V^{-1}Y$ and S a subspace spanned by the rows of A. Suppose that $S \cap N(X) = \{0\}$. Then MY is the S-BLUE of $f_S = E[M]_S$.

PROOF. From Prop. 3.1 (ii) the S-BLUE of f_S has the form $MY = A_1X'V^{-1}Y$ and we have $A_1T\beta = AT\beta = E_\beta[M]$ for all $\beta \in S$. Since

$S \cap N(X) = \{0\}$ we have that T_S (the restriction of T to S) induces an inner product T_S on S. Then $(A_1 - A)T\beta = 0$ for all $\beta \in S$ implies that each row of $A_1 - A$ is T_S-orthogonal to all the vectors in S, since the rows of $A_1 - A$ are in S, this implies that they are zero or equivalently $A_1 = A$.

As in the previous chapter a general version of the Gauss Markov Theorem can be obtained.

THEOREM 3.1.

Let a linear model $(X, V, \mathbb{R}^p)$. Let $MY \in L$ such that $MY = AX'V^{-1}Y$. Let S a subspace of $\mathbb{R}^p$ such that $S \cap N(X) = \{0\}$. Then MY is the unique S-BLUE of $E[M]_S$ if and only if there exists a matrix M_0 with rows in N(X) such that the rows of $A - M_0$ are in S.

If $v \notin N(X)$ we have that $v'Tv \neq 0$ and then $M_vY = \frac{1}{v'Tv} vv'X'V^{-1}Y$ is well defined. It is immediate that if $\beta = \lambda v$ then $E_\beta[M_v] = \lambda v = \beta$ and then also from Prop. 3.2 we can conclude that M_vY is the S-BLUE for $S = \mathrm{Span}\{v\}$. If $S \cap N(X) = \{0\}$ and $\dim S = r$, there exist a T_S-orthonormal basis $\{\beta_i\}_{i=1}^r$ for S. Using the T_S-orthonormality it is easy to see that:

$$E_\beta \left[\sum_{i=1}^{r} \beta_i \beta_i' X'V^{-1} \right] = \beta , \qquad \forall \beta \in S.$$

Again from Prop. 3.2 we can conclude that the S-BLUE $\hat{\beta}^S$ is given by

$$\hat{\beta}^S = \sum_{i=1}^{r} \beta_i \beta_i' X'V^{-1}Y.$$

THEOREM 3.2.

Let S a subspace of $\mathbb{R}^p$ such that $S \cap N(X) = \{0\}$. Then there exists an S-BLUE $\hat{\beta}^S$ and it is given by

$$\hat{\beta}^S = \sum_{i=1}^{r} \beta_i \beta_i' X' V^{-1} Y,$$

where $\{\beta_i\}_{i=1}^r$ is a T_S-orthonormal basis of S.

THEOREM 3.3.

Let S a subspace of $\mathbb{R}^p$. Then:

There exists an S-BLUE if and only if $S \cap N(X) = \{0\}$.

PROOF. Theorem 3.2 gives the "if" part. To see the other implication let $MY = \hat{\beta}^S$, then we have from Definition 2.2' that $E_\beta[\hat{\beta}^S] = \beta$ for all $\beta \in S$. We also have $E_\beta[M] = MX\beta = 0$ if $\beta \in N(X)$. Then if there exists $\beta \in S \cap N(X)$ different from zero we would have simultaneously $E_\beta[\hat{\beta}^S] = \beta$ and $E_\beta[\hat{\beta}^S] = 0$, which is impossible.

Among the S-BLUE's, when X is non full rank, there are some which are particularly important, those which are maximal in the sense that $\mathrm{Span}(S \cup N(X)) = \mathbb{R}^p$.

3.2. g-INVERSES AND S-BLUE's

The normal equations for the linear model $(X,V,\mathbb{R}^P)$ are given by

$$X'V^{-1}X\beta = X'V^{-1}Y \quad .$$

When X is a full rank matrix the solution is readily obtained using the inverse of $X'V^{-1}X$. When X is non full rank, $X'V^{-1}X$ is not invertible. The solution of the problem is now an affine variety and not just a point. In most practical situations, statisticians are not interested in the whole variety but mainly in obtaining a point in this variety. Two main methods have been used to obtain such a "solution point":

(i) Through the use of g-inverses of T.

(ii) Putting some restrictions on β in such a way that there is only one point in the variety which satisfies the restrictions.

The second method has the advantage over the first that the restrictions can have some meaning (or can even be "true") in the particular context where the problem arises. The first method has the "advantage" that no particular assumptions are made on β to obtain the solution, but the use of one particular g-inverse leads to an implicit choice of where β lies. It is now a well known fact that both approaches are equivalent in the sense that for each g-inverse there exists a set of linear restrictions such that the solution to the normal equations obtained by using them is equal to the one obtained by using the g-inverse. Also for each set of linear restrictions which define a unique solution to the normal equations there is a g-inverse with which the same solution is obtained.

Due mainly to their use for obtaining a solution to the normal equations, g-inverses have been widely studied. Geometric properties of

AA^-, A^-A and A^- (where A^- is a g-inverse of A) have been given (see for instance Rao 1962, Kruskal 1975) and also many methods for calculating them are available (see Shinozaki, 1975). Nevertheless the implications of the geometrical properties do not seem to have been fully exploited to give a method of calculating the g-inverses of T. The next theorem which is not essentially new is given mainly for the sake of completeness. We will first give some preliminary definitions and results. Given an arbitrary p×p matrix A, a g-inverse A^- of A is a matrix which satisfies

$$AA^-A = A \quad . \tag{3.2.1}$$

It is easy to see that any matrix which satisfies (3.2.1) must satisfy

$$CS(A^-) \cap N(A) = \{0\} \quad ,$$

and

$$\text{Span}\{CS(A^-A) \cup N(A)\} = \mathbb{R}^p ,$$

where $CS(A^-)$ is the vector space spanned by the columns of A^-. This suggests the following equivalence relation on the set $G(A)$ of g-inverses of A

$$A_1^- \sim A_2^- \quad \text{if and only if} \quad CS(A_1^-A) = CS(A_2^-A).$$

Given a subspace S of $\mathbb{R}^p$ such that $S \cap N(A) = \{0\}$ and Span $(S \cup N(A)) = \mathbb{R}^p$ $G_S(A)$ will be the equivalence class in $G(A)$ defined by the subspace S.

THEOREM 3.4.

Let S a subspace of $\mathbb{R}^p$ such that $S \cap N(X) = \{0\}$ and $\text{Span}(S \cup N(X)) = \mathbb{R}^p$. Let $G_S(T)$ the equivalence class of g-inverses of T defined by S. Let $\{\beta_i\}_{i=1}^r$ a T_S-orthonormal basis of S. Then:

(i) $\sum_{i=1}^{r} \beta_i\beta_i' \in G_S(T)$.

(ii) For any $T^- \in G_S(T)$ we have

$$\left(\sum_{i=1}^{r} \beta_i\beta_i'\right)T\beta = T^-T\beta, \qquad \text{for all } \beta \in \mathbb{R}^p.$$

PROOF. Let $\{u_j\}_{i=1}^{p-r}$ a basis for N(X), then:

$$T\left(\sum_{i=1}^{r} \beta_i\beta_i'\right)T\beta_i = T\beta_i, \qquad i = 1,\ldots,r.$$

$$T\left(\sum_{i=1}^{r} \beta_i\beta_i'\right)Tu_j = Tu_j = 0, \qquad j = 1,\ldots,p-r.$$

Since $\{\beta_i\}_{i=1}^{r} \cup \{u_j\}_{j=1}^{p-r}$ forms a basis of $\mathbb{R}^p$ this proves that $\sum_{i=1}^{r} \beta_i\beta_i'$ is a g-inverse of T. It is immediate that Range$\left(\sum_{i=1}^{r} \beta_i\beta_i'T\right) = S$. Then we have proved (i). To see (ii) notice that $TT^-T\beta_i = T\beta_i$, this implies that $T^-T\beta_i = \beta_i+v_i$ where $v_i \in N(X)$. Now if $v_i \neq 0$, since Range $(T^-T) = S$ we would have that $\beta_i+v_i-\beta_i \in S$ and so $S \cap N(X) \neq \{0\}$, contrary to our assumptions. Then:

$$\sum_{i=1}^{r} \beta_i\beta_i'T\beta_i = T^-T\beta_i, \qquad i = 1,\ldots,r,$$

and

$$\sum_{i=1}^{r} \beta_i\beta_i'Tu_j = T^-Tu_j = 0, \quad j = 1,\ldots,p-r,$$

if $T^- \in G_S(T)$. This implies (ii).

COROLLARY 3.4.1.

Let T^- a g-inverse of T such that Range $T^-T = S$. Then:

$$T^-X'V^{-1}Y = \hat{\beta}^S.$$

A very common situation in practical applications is to impose a set of linear restrictions: $H\beta = 0$. The theorem suggests the following method to obtain a solution which satisfies the restrictions.

METHOD. Let $S_0 = CS(H')$; applying Gram-Schmidt orthonormalization method to the rows of H an orthonormal basis $\{u_i\}_{i=1}^n$ of S_0 is obtained. Let $P_0 = \sum_{i=1}^n u_iu_i'$ the orthogonal projection of $\mathbb{R}^p$ onto S_0 and $\{e_i\}_{i=1}^p$ the canonical basis of $\mathbb{R}^p$. Let $v_i = e_i - P_0e_i$, $i = 1,\ldots,p$. Then it is immediate that: $\{v_i\}_{i=1}^p \perp S_0$ and $\text{Span}(\{v_i\}_{i=1}^p) = S_0^\perp$. It is worth noticing that $S_0^\perp = \{\beta | H\beta = 0\}$. We now apply to $\{v_i\}_{i=1}^p$ the Gram-Schmidt orthonormalization method (but using as "inner product" the one given by T) to obtain a maximal set $\{\beta_i\}_{i=1}^r$ of "T-orthonormal" vectors contained in $S_0^\perp$. Then the estimator $M_HY = \sum_{i=1}^r \beta_i\beta_i'X'V^{-1}Y$ satisfies the restrictions and is an S-BLUE (notice that $r = \dim S \leq \dim S_0^\perp$).

3.3. CHARACTERIZATION OF ALE WHEN X IS NON FULL RANK

Although the characterization that will be given at the end of this section is the one that would be intuitively expected from Theorem 2.3, Theorem 3.2 and Theorem 3.3, the formal proof is not immediate. We will need to generalize some results due to Shinozaki 1975 and Rao 1976. Those results will be useful also in other chapters.

PROPOSITION 3.3.

Let Q a n.n.d. pxp matrix. Let C a subset of $\mathbb{R}^p$. Let MY $\in$ L. If there exists an $M_0Y \in L$ and $\gamma \in C$ such that:

$$R_Q[M_0,\beta] \leq R_Q[M,\beta] \quad , \qquad \forall \ \beta \in C,$$

and

$$R_Q[M_0,\gamma] < R_Q[M,\gamma] \quad ,$$

there exists an NY $\in L$ such that:

$$R_I[N,\beta] \leq R_I[M,\beta] \quad , \qquad \forall \ \beta \in C,$$

and

$$R_I[N,\gamma] < R_I[M,\gamma] \quad .$$

PROOF. M_0 exists only if $Q \neq 0$. Then let $F = \frac{1}{\lambda} Q$ where λ is the largest eigenvalue of Q. Then

$$R_F[M_0,\beta] \leq R_F[M,\beta] \quad , \qquad \forall \ \beta \in C \ , \tag{3.3.1}$$

and

$$R_F[M_0,\gamma] < R_F[M,\gamma] \quad . \tag{3.3.2}$$

Let $NY = MY + F[M_0Y - MY]$. Then:

$$R_I[N,\beta] = R_I[M,\beta] + E_\beta[(M_0-M)'F^2(M_0-M)] + 2E_\beta[(M-\beta)'F(M_0-M)].$$

Since $0 \leq F \leq I$, we have $F^2 \leq F$ and

$$E_\beta[(M_0-M)'F^2(M_0-M)] \leq E_\beta[(M_0-M)'F(M_0-M)] \ ;$$

Also:

$$E_\beta[(M_0-M)'F(M_0-M)] = R_F[M_0,\beta] + R_F[M,\beta] - 2E_\beta[(M_0-\beta)'F(M-\beta)].$$

We then have

$$R_I[N,\beta] \leq R_I[M,\beta] + R_F[M_0,\beta] - R_F[M,\beta] \quad ,\forall \beta \in C.$$

From (3.3.1)

$$R_I[N,\beta] \leq R_I[M,\beta] \quad , \quad \forall \beta \in C.$$

From (3.3.2)

$$R_I[N,\gamma] < R_I[M,\gamma] \quad .$$

This proves the proposition.

By MY is an ALE under the risk Q, we will understand that MY satisfies the conditions obtained by substituting in Definition 2.1. R_I by R_Q. We have then the following known result (Shinozaki 1975, Rao 1976).

PROPOSITION 3.4.

Let Q a n.n.d p×p matrix. Let MY be an ALE, then MY is an ALE under the risk Q. (X full or non full rank).

In the next lines we will describe a generalization of unitary transformations and some related topics. As it is well known an unitary transformation from $\mathbb{R}^p$ into $\mathbb{R}^p$ is an isomorphism which transforms orthonormal sets into orthonormal sets. Given an arbitrary n.n.d.s. transformation T, the restriction T_S of T to a subspace S such that $S \cap N(X) = \{0\}$, defines an inner product on S. We will speak of T_S-unitary transformations as those linear transformations from S to S which transform T_S-orthonormal sets into T_S-orthonormal sets. If $B_1 = \{\beta_i\}_{i=1}^r$ and $B_2 = \{\gamma_i\}_{i=1}^r$ are two T_S-orthonormal bases of S, a matrix for the T_S-unitary transformation which transform B_1 into B_2 is given by

$$U = \left(\sum_{i=1}^{r} \gamma_i \beta_i'\right) T.$$

It is easy to see that if dim S = p - dim N(X)

$$U'T\,U = T .$$

This is a generalization of the property (for unitary transformations)

$$U'U = I.$$

Also any $p \times p$ matrix D such that Range $D = N(D)^{\perp} \subset S =$ Range T has a T_S-singular decomposition; i.e. there exist two T_S-unitary matrices U_1 and U_2 and a symmetric matrix Λ such that

$$D = U_2' \,\Lambda\, U_1 .$$

To see this first notice that $D'T^-D$ and DT^-D' are symmetric matrices. Here $T^- = \sum_{i=1}^{r} \beta_i\beta_i'$, $\{\beta_i\}_{i=1}^r$ is a T_S-orthonormal basis of S. If there exists a T_S-singular decomposition we should have

$$D'T^{-}D = U_1' \Lambda U_2 T^{-} U_2' \Lambda U_1 . \tag{3.3.3}$$

$$DT^{-}D' = U_2' \Lambda U_1 T^{-} U_1' \Lambda U_2 . \tag{3.3.4}$$

Now from Lemma 2.1 we know that there exist T_S-orthonormal bases $\{\gamma_i\}_{i=1}^r$, $\{\alpha_i\}_{i=1}^r$ of S such that

$$D'T^{-}D = T\,(\sum_{i=1}^{r} \lambda_i^2 \gamma_i \gamma_i')T,$$

$$DT^{-}D' = T\,(\sum_{i=1}^{r} \mu_i^2 \alpha_i \alpha_i')T,$$

(we chose λ_i^2 and μ_i^2 only for notation conveniences). From considerations arising from the forms of (3.3.3) and (3.3.4) and from the T_S-orthonormality of the γ_i's and α_i's, we must have that

$$U_1 = \sum_{i=1}^{r} \beta_i \gamma_i' T,$$

$$U_2 = \sum_{i=1}^{r} \beta_i \alpha_i' T,$$

$$\lambda_i^2 = \mu_i^2 , \qquad i = 1,\ldots,p,$$

and

$$\Lambda = T(\sum_{i=1}^{r} \lambda_i \beta_i \beta_i')T,$$

with $\{\beta_i\}_{i=1}^r$ a T_S-orthonormal basis of S. It is easy to check that those matrices satisfy (3.3.3) and (3.3.4). We are now ready to prove the next proposition.

PROPOSITION 3.5.

Let C a subset of $\mathbb{R}^p$. Let $MY = AX'V^{-1}Y$. Then if TAT is not a symmetric matrix there exists an estimator M_0Y such that:

$$R_T[M_0,\beta] < R_T[M,\beta] \quad , \qquad \forall\ \beta \in C.$$

PROOF. For any g-inverse T^- of T, we have

$$B_T[M,\beta] = -\beta'(I-TA')T(AT-I)\beta \quad ,$$

$$= -\beta'(T-TA'T)T^-(TAT-T)\beta \quad .$$

Let $D = T-TAT$, then TAT is symmetric if and only if D is symmetric. Let U_1, U_2 and Λ such that $D = U_2'\ \Lambda\ U_1$ is a T_0-singular decomposition of D (T_0 is the restriction of T to $N(X)^\perp$) defined as above. From this, if $\beta = \sum_{i=1}^{r} b_i\gamma_i + v$, $v \in N(X)$

$$B_T[M,\beta] = \sum_{i=1}^{r} \lambda_i^2 b_i^2 \quad .$$

We also have

$$\mathrm{Var}_T[M] = \mathrm{Tr}[TATA'] = \mathrm{Tr}[T^-TATT^-TA'T],$$

$$= \mathrm{Tr}[T^-(D-T)T^-(D'-T)],$$

$$= \mathrm{Tr}[T^-(DT^- - TT^-)(D'-T)],$$

and

$$= \mathrm{Tr}[T^-[DT^-D' - TT^-D' - DT^-T + TT^-T]] \quad .$$

Using the T_0-singular decomposition of D, we have

$$TT^{-}D' = D',$$

$$DT^{-}T = D,$$

$$DT^{-}D' = T(\sum_{i=1}^{r} \lambda_i^2 \alpha_i \alpha_i')T,$$

$$D' = U_1' \Lambda U_2 = T(\sum_{i=1}^{r} \lambda_i \gamma_i \alpha_i')T,$$

and

$$D = T(\sum_{i=1}^{r} \lambda_i \alpha_i' \gamma_i) T.$$

Then

$$\mathrm{Var}_T[M] = \sum_{i=1}^{r} \alpha_i'(DT^{-}D' - D' - D + T)\alpha_i,$$

$$= \sum_{i=1}^{r} (\lambda_i^2 - 2\lambda_i \alpha_i' T\gamma_i + 1).$$

Let

$$M_0 Y = (\sum_{i=1}^{r} (1-\lambda_i)\gamma_i\gamma_i')X'V^{-1}Y, \quad \text{then}$$

$$B_T[M_0,\beta] = \sum_{i=1}^{r} \lambda_i^2 b_i^2 = B_T[M,\beta] \quad ,$$

and

$$\mathrm{Var}_T[M_0] = \sum_{i=1}^{r} \lambda_i^2 - 2\lambda_i + 1.$$

If TAT is not symmetric, $U_1 \neq U_2$ and then there exist $1 \leq j \leq r$, such that $\alpha_j \neq \gamma_j$ and therefore $|\alpha_j' T\gamma_j| < 1$. Since for $1 \leq i \leq r$, $|\alpha_i' T\gamma_i| \leq 1$ we have that

$$\mathrm{Var}_T[M_0] < \mathrm{Var}_T\ [M].$$

Since $B_T[M_0,\beta] = B_T[M,\beta]$ for all $\beta \in \mathbb{R}^p$, this proves the proposition.

PROPOSITION 3.6.

Let C a subset of $\mathbb{R}^p$. Let $MY = AX'V^{-1}Y$. Suppose that $TAT = T(\sum_{i=1}^{r} (1-\lambda_i)\gamma_i\gamma_i')T$ and $\{\gamma_i\}_{i=1}^{r}$ is a T_0-orthonormal basis of $N(T)^{\perp}$. Then if $\lambda_j \notin [0,1]$ for some $1 \leq j \leq r$, and $\text{Span}(C) \supset \text{Span}\{\gamma_i\}_{(1-\lambda_i)\neq 0}$ there exists $M_0Y \in \mathcal{L}$ and $\gamma \in C$ such that:
$R_T[M_0,\beta] \leq R_T[M,\beta]$, $\forall \beta \in C$ and $R_T[M_0,\gamma] < R_T[M,\gamma]$.

PROOF. Let $M_0Y = A_0X'V^{-1}Y$ with $A_0 = \sum_{i=1}^{r} (1-\mu_i)\gamma_i\gamma_i'$ and $1-\mu_i = \min(1,|1-\lambda_i|)$. Let $\beta = \sum_{i=1}^{r} b_i\gamma_i + v$, $v \in N(T)$, then

$$B_T[M_0,\beta] = \sum_{i=1}^{r} \mu_i^2 b_i^2 ,$$

and

$$\text{Var}_T[M_0] = \sum_{i=1}^{r} (1-\mu_i)^2.$$

From the definition of μ_i it can be seen that $\lambda_i \notin [0,1]$ implies that $0 \leq \mu_i < |\lambda_i|$ and $\lambda_i \in [0,1]$ implies that $\mu_i = \lambda_i$. From this if $\lambda_j \notin [0,1]$ for some $1 \leq j \leq r$, we have that

$$B_T[M_0,\beta] = \sum_{i=1}^{r} \mu_i^2 b_i^2 \leq \sum_{i=1}^{r} \lambda_i^2 b_i^2 = B_T[M,\beta] \quad ,\forall \beta \in \mathbb{R}^p .$$

$$\text{Var}_T[M_0] = \sum_{i=1}^{r} (1-\mu_i)^2 < \sum_{i=1}^{r} (1-\lambda_i)^2 = \text{Var}_T[M].$$

and from the condition on C, there exists $\gamma \in C$ such that $\gamma T \gamma_j \neq 0$ and so $B_T[M_0, \gamma] < B_T[M,\gamma]$. This proves the proposition.

PROPOSITION 3.7.

If MY is an ALE, then $MY = AX'V^{-1}Y$ with TAT a symmetric matrix, $TAT = T(\sum_{i=1}^{r} \delta_i\beta_i\beta_i')T$, $0 \leq \delta_i \leq 1$ for $1 \leq i \leq r$ and $\{\beta_i\}_{i=1}^{r}$ is a T_0-orthonormal basis of $S_0 = N(T)^{\perp}$.

PROOF. As in Theorem 2.3, it is easy to see that any ALE must be the BLUE of its own expectation function, hence it must have the form $MY = AX'V^{-1}Y$ for some matrix A. From Prop. 3.4 MY is also an ALE under the risk defined by T, but then from Prop. 3.5 TAT must be symmetric; from Lemma 2.1 (b) $TAT = T(\sum_{i=1}^{r} \delta_i\beta_i\beta_i')T$ for a T_0-orthonormal basis $\{\beta_i\}_{i=1}^{r}$ of $S_0 = N(T)^{\perp}$. Finally from Prop. 3.6, $0 \leq \delta_i \leq 1$ for $1 \leq i \leq r$.

From the last proposition the condition that TAT is symmetric becomes important. Next we will characterize the set of matrices A which satisfy this condition. If $TAT = T(\sum_{i=1}^{r} \lambda_i\beta_i\beta_i')T$ with $\{\beta_i\}_{i=1}^{r}$ a T_0-orthonormal basis of $S_0 = N(T)^{\perp}$ we have that $TAT\beta_i = T\lambda_i\beta_i$, $1 \leq i \leq r$. This implies that $AT\beta_i = \lambda_i(\beta_i+\alpha_i)$ for some $\alpha_i \in N(T)$, $1 \leq i \leq r$. Let $\gamma_i = \beta_i+\alpha_i$, $1 \leq i \leq r$, then

$$\sum_{i=1}^{r} \lambda_i\gamma_i\gamma_i'T\beta = AT\beta , \qquad \forall\ \beta \in \mathbb{R}^p .$$

Let $A_1 = \sum_{i=1}^{r} \lambda_i\gamma_i\gamma_i'$. We will see that $A_2T = AT$ if and only if $A_2 = A_1+M_1$ where the rows of M_1 are in N(T). If one row of $M_1 \notin N(T)$ then for some β, $M_1T\beta \neq 0$ and so $A_2T\beta = (A_1T+M_1T)\beta \neq A_1T\beta = AT\beta$. The other implication is trivial. We will see now that $\{\gamma_i\}_{i=1}^{r}$ is a set of T_S-orthonormal vectors and $S = \mathrm{Span}\{\gamma_i\}_{i=1}^{r}$ satisfies $S \cap N(T) = \{0\}$.

$$\gamma_i'T\gamma_j = (\beta_i'+\alpha_i')T(\beta_j+\alpha_j) = \beta_i'T\beta_j = \delta_{ij} .$$

This proves the T_S-orthonormality. Suppose that $v = \sum_{i=1}^{r} a_i\gamma_i \in N(T)$, then $0 = \beta_i'Tv = a_i$, $1 \le i \le r$. This implies that $v = 0$ and so $S \cap N(T) = \{0\}$. We obtain the next proposition.

PROPOSITION 3.8.

TAT is symmetric if and only if $A = \sum_{i=1}^{r} \lambda_i\gamma_i\gamma_i' + M_1$ where $\{\gamma_i\}_{i=1}^{r}$ is a T_S-orthonormal basis of S = Range AT and the rows of M_1 are in $N(T)$.

Now we will give the main result of this section.

THEOREM 3.5.

Let $MY \in L$, S = Range M. Then MY is an ALE if and only if $S \cap N(X) = \{0\}$ and there exists a matrix A such that $MY = AX'V^{-1}Y$, $A = \sum_{i=1}^{r} \delta_i\gamma_i\gamma_i'$, $\{\gamma_i\}_{i=1}^{r}$ is a T_S-orthonormal basis of S and $0 \le \delta_i \le 1$ for $1 \le i \le r$.

PROOF. The necessity of this condition is a direct consequence of Prop. 3.7 and Prop. 3.8. To see the sufficiency let us suppose that there exists $M_0Y \in L$ such that $M_0Y = A_0X'V^{-1}Y$ and

$$R_I[M_0,\beta] \le R_I[M,\beta] \quad , \qquad \forall\ \beta \in \mathbb{R}^p .$$

This implies that

$$R_I[M_0,\beta] \le R_I[M,\beta] \quad , \qquad \forall\ \beta \in S. \tag{3.3.5}$$

There is no loss of generality in taking the rows of A_0 in S. Let X_1 such that $X+X_1$ is a full rank matrix and $P_SX_1' = 0$ (P_S is the orthogonal projection onto S). Let $NY = A(X+X_1)'V^{-1}Y$ and $N_0Y = A_0(X+X_1)'V^{-1}Y$. We have that

$$R_I[N,\beta] = R_I[M,\beta] \quad , \qquad \forall\ \beta \in S.$$

$$R_I[N_0,\beta] = R_I[M_0,\beta] \quad , \qquad \forall\ \beta \in S.$$

(Because rows of A and A_0 are in S and $P_S X_1' = 0$). From Theorem 3.6, NY is admissible for the model $(X+X_1, V, S)$ and therefore there does not exist $N_0Y \in L$ and $\beta_0 \in S$ such that

$$R_I[N_0,\beta] \leq R_I[N,\beta] \quad , \qquad \forall \beta \in S,$$

and

$$R_I[N_0,\beta_0] < R_I[N,\beta_0] \quad .$$

(3.3.5) implies that

$$R_I[N_0,\beta] = R_I[N,\beta] \quad , \qquad \forall \beta \in S . \tag{3.3.6}$$

Since NY is ALE for the model $(X+X_1, V, S)$, (3.3.6) implies that N_0Y is also an ALE for $(X+X_1, V, S)$. But this implies, again by Theorem 3.6, that A_0 is symmetric. Let $T_S^- = \sum_{i=1}^{r} \gamma_i\gamma_i'$ then for $\beta \in S$: $T_S^- T\beta = I_S\beta = \beta$. Hence

$$B_I[M,\beta] = \beta'T(A-T_S^-)(A-T_S^-)T\beta \quad , \quad \forall \beta \in S,$$

and

$$B_I[M_0,\beta] = \beta'T(A_0-T_S^-)(A_0-T_S^-)T\beta, \quad \forall \beta \in S.$$

Let $\gamma = T\beta$. Then from (3.3.6) we have

$$\gamma'\pi\gamma = \gamma'[(A-T_S^-)^2 - (A_0-T_S^-)^2]\gamma = 0, \quad \forall \gamma \in \text{Range } T.$$

Since π is symmetric, the rows of π are in S and $S \cap N(T) = \{0\}$

$$(A-T_S^-)^2 - (A_0-T_S^-)^2 = 0.$$

Since $T_S^- - A$ and $T_S^- - A_0$ are symmetric and n.n.d. this implies that $A = A_0$. As a consequence we obtain.

COROLLARY 3.5.1.

$MY = AX'V^{-1}Y$ is an ALE if and only if TAT is symmetric and $(TAT)T^-(TAT) \leq TAT$. (T^- is any g-inverse of T).

PROOF. If MY is an ALE from Prop. 3.7 TAT is a symmetric matrix and $TAT = T(\sum_{i=1}^{r} \delta_i \beta_i \beta_i')T$, $0 \leq \delta_i \leq 1$ for $1 \leq i \leq r$.

Then $(TAT)T^-(TAT) = T(\sum_{i=1}^{r} \delta_i^2 \beta_i \beta_i')T$. From this it is immediate that $(TAT)T^-(TAT) \leq TAT$. To see the other implication; since TAT is symmetric from Lemma 2.1 (b) $TAT = T(\sum_{i=1}^{r} \lambda_i \beta_i \beta_i')T$ for some T_0-orthonormal basis $\{\beta_i\}_{i=1}^{r}$ (T_0 is the restriction of T to $N(T)^{\perp}$). From $(TAT)T^-(TAT) \leq TAT$ it follows that $\lambda_i^2 \leq \lambda_i$ and this implies that $0 \leq \lambda_i \leq 1$ for $1 \leq i \leq r$. Prop. 3.8 and Theorem 3.5 imply that MY is an ALE. This corollary generalizes to the non full rank case the characterization given by Rao 1976, for ALE's in the full rank case (see Corollary 2.3.1).

COROLLARY 3.5.2.

The ridge estimator $(T+kI)^{-1}X'V^{-1}Y$ is an ALE even if X is a non full rank matrix.

PROOF. $(T+kI)^{-1} = A+M_1$ where A and M_1 are symmetric matrices, rows of A are in $N(T)^{\perp}$ and rows of M_1 are in N(T). Then $(T+kI)^{-1}X'V^{-1} = AX'V^{-1}$. It is also easy to check that $A = \sum_{i=1}^{r} \delta_i \beta_i \beta_i'$ where $0 \leq \delta_i \leq 1$ for $1 \leq i \leq r$ and $\{\beta_i\}_{i=1}^{r}$ is a T_0-orthonormal basis of $N(T)^{\perp}$. From Theorem 3.5 we can conclude that $(T+kI)^{-1}X'V^{-1}Y$ is an ALE.

3.4. LINEAR ADMISSIBLE ESTIMATORS FOR SUBSPACES

In the thesis, until now, we have studied admissible linear estimators when the parameter space is $\mathbb{R}^p$. In this section the notion of linear admissibility, when the parameter space is a subspace of $\mathbb{R}^p$, will be studied. Hoffmann (1977), studied the problem for ellipsoids centred at the origin. First some propositions will be given. The next one, which is very useful, is essentially Lemma 1 in Hoffmann 1977.

PROPOSITION 3.9.

Let Λ and Q two p.d. symmetric pxp matrices. Let MY, NY $\in L$. Then there exists $M_0Y \in L$ such that

$$R_Q[M_0,\beta] \leq R_Q[M,\beta] \quad \text{for all } \beta \text{ such that } R_\Lambda[N,\beta] \leq R_\Lambda[M,\beta] \quad ,$$

and

$$R_Q[M_0,\beta] < R_Q[M,\beta] \quad \text{for all } \beta \text{ such that } R_\Lambda[N,\beta] < R_\Lambda[M,\beta] \quad .$$

PROOF. Let λ the maximum eigenvalue of Λ and μ the minimum eigenvalue of Q, let $D = \frac{\Lambda}{\lambda}$ and $\pi = \frac{Q}{\mu}$. Let

$$M_0 = (I-\pi^{-1}D)M + \pi^{-1}DN.$$

We then have

$$R_\pi[M_0,\beta] = E_\beta[(M_0-\beta)\pi(M_0-\beta)],$$

$$= E_\beta[(M-\beta)'\pi(M-\beta)] + E_\beta[(M-N)'D\pi^{-1}\pi\pi^{-1}D(M-N)]$$

$$-E_\beta[(M-N)'D\pi^{-1}\pi(M-\beta)] - E_\beta[(M-\beta)'\pi\pi^{-1}D(M-N)],$$

$$\leq R_\pi[M,\beta] + E_\beta[(M-N)'D(M-N)] - 2E_\beta[(M-N)'D(M-\beta)],$$

$$= R_\pi[M,\beta] + R_D[N,\beta] - R_D[M,\beta]$$

$$\leq R_\pi[M,\beta] \quad , \qquad \text{if } R_\Lambda[N,\beta] \leq R_\Lambda[M,\beta] \quad ,$$

or

$$< R_\pi[M,\beta] \quad , \quad \text{if } R_\Lambda[N,\beta] < R_\Lambda[M,\beta] \quad .$$

The first inequality is a consequence of $D\pi^{-1}D \leq D$. The last two inequalities are a consequence of the hypothesis of the proposition and the fact that $R_\Lambda[N,\beta] \leq R_\Lambda[M,\beta]$ if and only if $R_D[N,\beta] \leq R_D[M,\beta]$, (idem for $<$). The same reason and the inequalities obtained imply the proposition.

PROPOSITION 3.10.

Let Q be a p.d. symmetric p×p matrix. Let X full rank. If $MY \in L$ is not of the form $MY = AX'V^{-1}Y$, $A = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'$, $\{\beta_i\}_{i=1}^{r}$ a T-orthonormal basis of S and $0 \leq \delta_i \leq 1$ for $1 \leq i \leq r$, there exists an estimator $M_0Y \in L$ and $\gamma \in S$ such that

$$R_Q[M_0, \beta] \leq R_Q[M, \beta] \,, \forall \beta \in S \quad \text{and } R_Q[M_0,\gamma] < R_Q[M,\gamma] \;. \qquad (3.4.1)$$

PROOF. If MY is defined as above, Propositions 3.5 and 3.6 show the existence of an estimator $NY \in L$ and $\gamma \in S$

$$R_T[N,\beta] \leq R_T[M,\beta], \forall \beta \in S \text{ and } R_T[N,\gamma] < R_T[M,\gamma] \;.$$

Since X is full rank, T is invertible and so p.d. Making $T = \Lambda$, (3.4.1) follows from Proposition 3.9. This proves the proposition.

PROPOSITION 3.11.

Let Q be a p.d. symmetric p×p matrix. Let X full rank. Then MY is an ALE if and only if MY is an ALE under the risk defined by Q.

PROOF. The proposition is an immediate consequence of Proposition 3.9, the definition of ALE and ALE under the risk given by Q.

The last proposition was already proved by Shinozaki (1975). The next definition makes precise the notion of linear admissible estimators for subspaces.

DEFINITION 3.1.

Let Q a p.d. symmetric $p \times p$ matrix and S a subspace of $\mathbb{R}^p$. Let $MY \in L$. Then MY is a S-ALE under the risk Q if and only if there does not exist $NY \in L$ and $\gamma \in S$ such that

$$R_Q[N,\beta] \leq R_Q[M,\beta] \quad , \quad \forall \beta \in S$$

and

$$R_Q[N,\gamma] < R_Q[M,\gamma] \ .$$

When Q = I we will simply say that MY is an S-ALE.

Proposition 3.9 implies that MY is an S-ALE under Q if and only if MY is an S-ALE. The next proposition is important for the main result of the section.

PROPOSITION 3.12.

Let S a subspace of $\mathbb{R}^p$. Let $MY \in L$. Then

(i) If Range $M \not\subset S$, $R_Q[P_S^Q M,\beta] < R_Q[M,\beta]$, for all $\beta \in S$.

(ii) If MY is an S-ALE then Range $M \subset S$.

PROOF. For any $\beta \in S$ we have

$$B_Q[M,\beta] = \|(MX-I)\beta\|_Q^2 \geq \|P_S^Q(MX-I)\beta\|_Q^2 ,$$

$$= \|(P_S^Q MX-I)\beta\|_Q^2 = B_Q[P_S^Q M,\beta] \quad . \qquad (3.4.2)$$

Let $S_0 = S^{\perp Q}$. Then

$$\mathrm{Var}_Q[M] = \mathrm{Tr}[QMVM'] = \mathrm{Tr}[(P_S^Q+P_{S_0}^Q)'Q(P_S^Q+P_{S_0}^Q)MVM'] ,$$

$$= \mathrm{Tr}[(P_S^Q)'Q(P_S^Q)MVM'] + \mathrm{Tr}[(P_{S_0}^Q)'Q(P_{S_0}^Q)MVM'] ,$$

$$= \mathrm{Tr}[QP_S^Q MVM'(P_S^Q)'] + \mathrm{Tr}[Q(P_{S_0}^Q)MVM'(P_{S_0}^Q)'] ,$$

$$= \mathrm{Var}_Q[P_S^Q M] + \mathrm{Var}_Q[P_{S_0}^Q M] .$$

If Range $M \not\subset S$ we have $\mathrm{Var}_Q[P_{S_0}^Q M] > 0$ and

$$\mathrm{Var}_Q[M] > \mathrm{Var}_Q[P_S^Q M] . \qquad (3.4.3)$$

From (2.2.2), (3.4.2) and (3.4.3) we have

$$R_Q[P_S^Q M,\beta] < R_Q[M,\beta] , \qquad \forall \beta \in S.$$

This proves (i). And (ii) is an immediate consequence of the definition of S-ALE and (i).

THEOREM 3.6.

Let $MY \in L$. Let S be a subspace of $\mathbb{R}^p$. Then

(i) MY is an S-ALE if and only if Range $M \subset S$ and MY is an ALE.

(ii) MY is an S-ALE if and only if $MY = \sum_{i=1}^{r} \delta_i \beta_i \beta_i' X'V^{-1}Y$ where $\{\beta_i\}_{i=1}^{r}$ is a T-orthonormal basis of S and $0 \le \delta_i \le 1$, $1 \le i \le r$.

PROOF. Proposition 3.10 implies that if MY is an S-ALE then it is an ALE. Proposition 3.12 implies that if MY is an S-ALE then Range $M \subset S$. This proves the first implication in (i). From Theorem 2.3 Range $M \subset S$ and MY is an ALE is equivalent to asking that MY has the form given in (ii). Then to prove the theorem we only need to see that any estimator MY of the form given in (ii) is an S-ALE. But this last implication follows from Theorem 4.1 and Theorem 4.2 using a similar argument as in Proposition 4.8, restricted to S.

CHAPTER 4

The chapter studies minimax estimators over some subsets of $\mathbb{R}^p$, when the risk considered is a quadratic risk defined by a p.d. symmetric matrix Q and attention is restricted to the set of linear estimators. Those estimators will be called Q-minimax estimators. We will assume that X is a full rank matrix.

The first section of the chapter studies a particular class of subsets of $\mathbb{R}^p$ which will be called T-cubes. This class appears to be in some sense a "natural" class of sets for the problem. It also contains a wide variety of sets which can make it useful for applications. As opposed to the second class of sets studied in the chapter, this class can be defined by linear restrictions on the parameter β.

The second section of the chapter studies the problem for the class of ellipsoids with center at the origin. The ellipsoids are defined by quadratic constraints on the parameter β. Part of the problem for this class has been already studied by Kuks (1972) and Lauter (1975).

4.1 Q-MINIMAX ESTIMATORS OVER T-CUBES

The section is structured as follows:
Proposition 4.1 shows that the Q-minimax estimator over a T-cube is the same as that over the set of vectors formed by its "corners". That is why the whole section studies Q-minimax estimators of T-orthogonal subsets C of $\mathbb{R}^p$. Proposition 4.2 simplifies considerably the study by relating the axis (see end of Chapter 2) of the Q-minimax estimator of C, to the vectors of C. The unicity of the Q-minimax estimator M_cY of C is proved in Proposition 4.3. From Proposition 4.2 we only need to determine the shrinkage factors (see end of Chapter 2) of M_cY to have it totally specified. A property satisfied by the bias of M_c on C (proved in Proposition 4.4) allows us to calculate, using Lagrange multipliers, a condition on the shrinkage factors which defines them implicitly. This is mainly the content of Proposition 4.5. Proposition 4.6 shows how to build a set C_M for which the ALE MY is a Q-minimax estimator, if the shrinkage factors of MY are less than unity (the case when some of the shrinkage factors are one can be treated using the results of Theorem 4.2). Proposition 4.7, translates the results of Proposition 4.6 into the usual representation for ALE's. Theorem 4.1 gives an explicit expression for the shrinkage factors of M_cY if it is known on which axis of M_cY the shrinkage factors will be different from zero. As a consequence of the theorem we are able to give a method to calculate the Q-minimax estimator of an arbitrary T-orthogonal set of vectors C. After this some particular examples are studied, where the previous results are used. Theorem 4.2 is essentially the natural extension of Theorem 4.1. Proposition 4.8, give us a result which is a consequence of the previous

results of the chapter and whose main application is in Theorem 2.2. The section ends with Proposition 4.9 and Theorem 4.3 which are the starting point for the second section.

The next two definitions make precise our terminology.

DEFINITION 4.1.

Let $MY \in L$. Let Z be any subset of $\mathbb{R}^p$. If

$$\sup_{\beta \in Z} R_Q[M,\beta] \leq \sup_{\beta \in Z} R_Q[N,\beta] \quad , \quad \forall\ NY \in L \ ,$$

we will say that MY is a Q-minimax estimator for Z. Sometimes Q-minimax estimator will be written Q-m.e.

DEFINITION 4.2.

Let Z be a subset of $\mathbb{R}^p$. Let cc(Z) denote the convex hull of Z and let $\mathrm{Sym}(Z) = \{\lambda\beta \mid \lambda \in [-1,1],\ \beta \in Z\}$. Then Z will be a T-cube if there exist a set C_z of T-orthogonal vectors of $\mathbb{R}^p$ such that

$$Z = cc(\mathrm{Sym}(C_z)).$$

Sometimes the elements of C_z will be called corners of Z.

The next proposition simplifies considerably the study of Q-minimax estimators for T-cubes.

PROPOSITION 4.1.

Let Z be a T-cube. Let C_z a set of corners of Z. Then MY is a Q-minimax estimator for C_z if and only if MY is a Q-minimax estimator for Z.

PROOF. For all $MY \in L$ we have

$$\sup_{\beta \in C_z} R_Q[M,\beta] = \sup_{\beta \in \mathrm{Sym}(C_z)} R_Q[M,\beta] = \sup_{\beta \in cc(\mathrm{Sym}(C_z))} R_Q[M,\beta]$$

$= \sup_{\beta \in Z} R_Q[M,\beta]$.

The second equality follows because $R_Q[M,\beta]$ is a convex function on β and the first equality from the definition of $R_Q[M,\beta]$ and $\mathrm{Sym}(C_z)$.

From the proposition it is clear that the problem of finding a Q-m.e. for a T-cube Z is the same as the one of finding a Q-m.e. for one of its sets of corners C_z; it is for this reason that from now on we will deal only with Q-m.e. for T-orthogonal subsets of $\mathbb{R}^p$.

PROPOSITION 4.2.

Let $C = \{\gamma_i\}_{i=1}^m$ be a subset of T-orthogonal vectors of $\mathbb{R}^p$. Let $\beta_i = \frac{\gamma_i}{\|\gamma_i\|_T}$, $(1 \leq i \leq m)$. Then any Q-minimax estimator MY for C has the form

$$MY = \sum_{i=1}^{m} \delta_i \beta_i \beta_i' X' V^{-1} Y ,$$

where δ_i depends on Q, T and C.

PROOF. Any Q-minimax estimator must be the BLUE of its expectation function, a similar argument to the one given at the beginning of Theorem 2.3 shows that $MY = AX'V^{-1}Y$. Now let P_i^Q the Q-orthogonal projection of $\mathbb{R}^p$ onto the $S_i = \mathrm{Span}\{\beta_i\}$. Let $M_0 Y = A_0 X'V^{-1}Y$ such that

$$A_0 T\gamma_i = P_i^Q AT\gamma_i, \quad \forall\ \gamma_i \in C,$$

and

$$A_0 T\gamma = 0, \quad \forall\ \gamma \in C^{\perp T}.$$

Since T is invertible M_0 is uniquely defined and since:

$$B_Q[M,\gamma_i] = \|(AT-I)\gamma_i\|_Q^2 \quad ,$$

and

$$B_Q[M_0,\gamma_i] = \| (P_i^Q\ AT-I)\gamma_i \|_Q^{\ 2} = \| P_i^Q\ (AT-I)\gamma_i \|_Q^{\ 2},$$

we have

$$B_Q[M,\gamma_i] \geq B_Q[M_0,\gamma_i] \quad , \qquad 1 \leq i \leq m\ . \tag{4.1.1}$$

Let $\{\beta_i\}_{i=1}^p$ a T-orthonormal basis of $\mathbb{R}^p$ obtained when $\{\beta_i\}_{i=1}^m$ is completed. Then from Lemma 2.1(f) we have

$$T = T(\sum_{i=1}^{p} \beta_i\beta_i')T.$$

Therefore

$$\begin{aligned} Var_Q[M] &= Tr[QATA'] = Tr[QAT(\sum_{i=1}^{p} \beta_i\beta_i')TA'] \quad , \\ &= \sum_{i=1}^{p} Tr[QAT\beta_i\beta_i'TA'] = \sum_{i=1}^{p} \| AT\beta_i \|_Q^{\ 2} \ , \end{aligned}$$

and

$$Var_Q[M_0] = \sum_{i=1}^{m} \| P_i^Q\ AT\beta_i \|_Q^{\ 2} \ .$$

Then

$$Var_Q[M] \geq Var_Q[M_0] \ , \tag{4.1.2}$$

with strict inequality if for some $1 \leq j \leq m$, $AT\beta_j$ is not in S_j or if $AT\beta_i \neq 0$ for $m < i \leq p$. From (2.2.2), adding (4.1.1) and (4.1.2) we obtain

$$R_Q[M,\gamma_i] \geq R_Q[M_0,\gamma_i] \quad , \qquad 1 \leq i \leq m,$$

with strict inequality if $Var_Q[M] > Var_Q[M_0]$. This implies that MY must be of the form stated in the proposition if it is a Q-minimax estimator of C.

The next proposition shows that for each T-orthogonal subset C of $\mathbb{R}^p$ there exists a unique Q-minimax estimator MY of C.

PROPOSITION 4.3.

Let $C = \{\gamma_i\}_{i=1}^m$ a set of T-orthogonal vectors of $\mathbb{R}^p$. Then there exists a unique Q-minimax estimator MY of C.

PROOF. From the previous proposition any Q-m.e must be of the form

$$M_{\underset{\sim}{\delta}} = \sum_{i=1}^{m} \delta_i \beta_i \beta_i' X'V^{-1}Y \quad \text{where} \quad \beta_i = \gamma_i / \| \gamma_i \|_T \quad \text{and} \quad \underset{\sim}{\delta} = (\delta_1, \dots \delta_m).$$

Then

$$R_Q[M_{\underset{\sim}{\delta}}, \gamma_i] = (1-\delta_i)^2 \| \gamma_i \|_Q^{\,2} + \sum_{i=1}^{m} \delta_i^{\,2} \| \beta_i \|_Q^{\,2} = f_i(\underset{\sim}{\delta}) .$$

Each $f_i(\underset{\sim}{\delta})$ is a quadratic function of $\underset{\sim}{\delta}$, so that each f_i has a unique minimum. Now the supremum of a finite number of functions which have each a unique minimum has itself a unique minimum. Thus there exist $\underset{\sim}{\delta}_0$ such that

$$\sup_{1 \le i \le m} f_i(\underset{\sim}{\delta}) > \sup_{1 \le i \le m} f_i(\underset{\sim}{\delta}_0), \quad \forall\, \underset{\sim}{\delta} \in \mathbb{R}^m \text{ such that } \underset{\sim}{\delta} \neq \underset{\sim}{\delta}_0 .$$

$$\sup_{1 \le i \le m} R_Q[M_{\underset{\sim}{\delta}}, \gamma_i] > \sup_{1 \le i \le m} R_Q[M_{\underset{\sim}{\delta}_0}, \gamma_i] \quad , \forall\, \underset{\sim}{\delta} \in \mathbb{R}^m, \quad \underset{\sim}{\delta} \neq \underset{\sim}{\delta}_0 .$$

This proves the proposition.

From now on M_cY will be the Q-minimax estimator of C. Also it can happen that some of the shrinkage factors of M_cY are zero. There is no loss of generality in assuming that these shrinkage factors correspond to the larger i's; that is we can always reorder the vectors γ_i in such a way that

$$M_cY = \sum_{i=1}^{r} \delta_i \beta_i \beta_i' X'V^{-1}Y + \sum_{i=r+1}^{m} \delta_i \beta_i \beta_i' X'V^{-1}Y = \sum_{i=1}^{r} \delta_i \beta_i \beta_i' X'V^{-1}Y,$$

$\delta_i \neq 0$ if $1 \leq i \leq r$ and $\delta_i = 0$ if $r < i \leq m$. This ordering assumption will apply from now on. The next proposition gives an important property satisfied by the bias of M_cY on C.

PROPOSITION 4.4.

Let $C = \{\gamma_i\}_{i=1}^m$ a set of T-orthogonal vectors of $\mathbb{R}^p$ and

$\beta_i = \dfrac{\gamma_i}{\|\gamma_i\|_T}$, $1 \leq i \leq m$. Let $M_cY = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y$ the Q-minimax estimator of C. Then

(a) $B_Q[M_c,\gamma_i] = B_Q[M_c,\gamma_j]$, $\quad 1 \leq i,\ j \leq r$.

(b) $B_Q[M_c,\gamma_i] \geq B_Q[M_c,\gamma_k]$, $\quad 1 \leq i \leq r, \quad r < k \leq m$.

(c) $B_Q[M_c,\gamma_i] = \sup_{\gamma\in C} B_Q[M_c,\gamma]$, $\quad 1 \leq i \leq r$.

PROOF. It is immediate that (c) implies (a). Let us then suppose that (c) does not hold and let n such that $1 \leq n \leq r$ and

$$B_Q[M_c,\gamma_n] < \sup_{\gamma\in C} B_Q[M_c,\gamma] \quad . \tag{4.1.3}$$

We will build an estimator M_0Y which is "better" than M_cY for C. Let

$M_0Y = \sum_{\substack{i=1\\ i\neq n}}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y + \delta\beta_n\beta_n'X'V^{-1}Y$. Then for any $\delta < \delta_i$ we have

$$\mathrm{Var}_Q[M_0] < \mathrm{Var}_Q[M_c] \quad , \tag{4.1.4}$$

and

$$B_Q[M_0,\gamma_i] = B_Q[M_c,\gamma_i] \quad , \quad 1 \leq i \leq m, \quad i \neq n.$$

Therefore

$$R_Q[M_0,\gamma_i] < R_Q[M_c,\gamma_i] \ , \quad 1 \leq i \leq m, \quad i \neq n. \tag{4.1.5}$$

As $\delta \to \delta_i$, $B_Q[M_0,\gamma_n] \to B_Q[M_c,\gamma_n]$ and then from (4.1.3) for $\delta < \delta_i$ close enough to δ_i we must have

$$B_Q[M_0,\gamma_n] < \sup_{\gamma\in C} B_Q[M_c,\gamma] \quad .$$

Thus from (4.1.4) we have

$$R_Q[M_0,\gamma_n] < \sup_{\gamma\in C} R_Q[M_c,\gamma] \quad .$$

From this last inequality and (4.1.5) we obtain

$$\sup_{\gamma\in C} R_Q[M_0,\gamma] < \sup_{\gamma\in C} R_Q[M_c,\gamma] \quad .$$

Then (4.1.3) is impossible and (c) is proved. To see (b) we again suppose that it does not hold and then show the existence of an estimator M_0Y which is "better" than M_cY for C. Let now n be such that

$n > r$ and $R_Q[M_c,\gamma_n] > \sup_{1\le i\le r} R_Q[M_c,\gamma_i]$.

And let $M_0Y = MY + \sum_{i=r+1}^{m} \delta\beta_i\beta_i'X'V^{-1}Y$. Then

$$R_Q[M_0,\gamma_i] = (1-\delta)^2\|\gamma_i\|_Q^2 + \delta^2 \sum_{i=r+1}^{m} \|\beta_i\|_Q^2 + \mathrm{Var}_Q[M_c],$$

$$r < i \le m. \qquad (4.1.6)$$

Now if $\delta > 0$ is chosen small enough we have

$$(1-\delta)^2\|\gamma_i\|_Q^2 + \delta^2 \sum_{i=r+1}^{m} \|\beta_i\|_Q^2 < \|\gamma_i\|_Q^2 \ , \ r < i \le m, \qquad (4.1.7)$$

and

$$R_Q[M_0,\gamma_i] = R_Q[M_c,\gamma_i] + \delta^2 \sum_{i=r+1}^{m} \|\beta_i\|_Q^2 < \sup_{\gamma\in C} R_Q[M_c,\gamma], \quad 1 \le i \le r. \qquad (4.1.8)$$

From (4.1.6) and (4.1.7) we obtain

$$R_Q[M_0,\gamma_i] < \| \gamma_i \|_Q^2 + Var_Q[M_c] = R_Q[M_c,\gamma_i] \quad , \quad r < i \leq m. \qquad (4.1.9)$$

Finally from (4.1.8) and (4.1.9) we have

$$\sup_{\gamma \in C} R_Q[M_0,\gamma] < \sup_{\gamma \in C} [M_c,\gamma] \quad ,$$

which contradicts the assumption on M_cY. The proposition is proved.

PROPOSITION 4.5.

Let $C = \{\gamma_i\}_{i=1}^m$ a set of T-orthogonal vectors of $\mathbb{R}^p$. Let $\beta_i = \frac{\gamma_i}{\| \gamma_i \|_T}$, $1 \leq i \leq m$. Let $M_cY = \sum_{i=1}^{r} \delta_i \beta_i \beta_i ' X'V^{-1}Y$ the Q-minimax estimator of C. Then M_cY must satisfy

(a) $0 < \delta_i < 1, \quad 1 \leq i \leq r.$

(b) $(1-\delta_i)\| \gamma_i \|_Q = (1-\delta_j)\| \gamma_j \|_Q \quad , \quad 1 \leq i, \; j \leq r.$

(c) $\sum_{i=1}^{r} \frac{\delta_i}{1-\delta_i} \frac{1}{\| \gamma_i \|_T^2} = 1.$

PROOF. From Prop 3.10 and 4.3, we know that $0 \leq \delta_i \leq 1$ for $1 \leq i \leq r$. By the ordering assumptions adopted just before Proposition 4.4, $\delta_i > 0$ for $1 \leq i \leq r$. Then, we only need to prove that $\delta_i \neq 1$ for $1 \leq i \leq r$. To see this suppose that $\delta_n = 1$ for some $1 \leq n \leq r$ and let $M_0Y = M_cY + M_nY - \beta_n\beta_n 'X'V^{-1}Y$, where M_nY is the MMSELE of γ_n. Then

$$Var_Q[M_0] = Var_Q[M_c] - \| \beta_n \|_Q^2 + Var_Q[M_n] < Var_Q[M_c].$$

$$B_Q[M_0,\gamma_i] = B_Q[M_c,\gamma_i] \quad , \quad 1 \leq i \leq m, \qquad i \neq n.$$

Joining the last two expressions we obtain

$$R_Q[M_0,\gamma_i] < R_Q[M_c,\gamma_i] \quad , 1 \leq i \leq m, \qquad i \neq n. \qquad (4.1.10)$$

Also, since $R_Q[M_n,\gamma_n] < \|\beta_n\|_Q^2$, we have

$$R_Q[M_0,\gamma_n] = R_Q[M_n,\gamma_n] + \text{Var}_Q[M_c] - \|\beta_n\|_Q^2 < \text{Var}_Q[M_c] = R_Q[M_c,\gamma_n].$$

From (4.1.10) and the last inequality we have

$$\sup_{\gamma\in C} R_Q[M_0,\gamma] < \sup_{\gamma\in C} R_Q[M_c,\gamma] \quad .$$

This contradicts that M_cY is the Q-m.e. for C. Then $\delta_i < 1$, $1 \leq i \leq r$ and (a) is proved. The condition (b) is essentially Proposition 4.4 (a) (if it is known that $0 < \delta_i < 1$). The problem of finding M_cY when it is known that M_cY has the form given in the proposition can be put in terms of a minimization problem with Lagrange multipliers if the restrictions on the bias proved in Proposition 4.4 are used. Since $R_Q[M_c,\gamma_i]$ as a function of $\underset{\sim}{\delta} = (\delta_1,\ldots\delta_r)$ is quadratic, the solution of the Lagrange problem is unique and it is a minimum. We have, if $1 \leq n \leq r$, for all $1 \leq i \leq r$

$$\frac{\partial}{\partial\delta_i} R_Q[M_c,\gamma_n] - \sum_{\substack{i=1\\i\neq n}}^{r} \mu_i[B_Q[M_c,\gamma_n] - B_Q[M_c,\gamma_i]] = 0.$$

Or

$$\frac{\partial}{\partial\delta_i} (1-\delta_n)^2\|\gamma_n\|_Q^2 + \sum_{i=1}^{r} \delta_i^2\|\beta_i\|_Q^2 - \sum_{\substack{i=1\\i\neq n}}^{r} \mu_i[(1-\delta_n)^2\|\gamma_n\|_Q^2 - (1-\delta_i)^2\|\gamma_i\|_Q^2] = 0 .$$

Then

$$-2(1-\delta_n)\|\gamma_n\|_Q^2 + 2\delta_n\|\beta_n\|_Q^2 + 2\sum_{\substack{i=1\\ i\neq n}}^{r} \mu_i[(1-\delta_n)\|\gamma_n\|_Q^2] = 0,$$

and

$$2\delta_i\|\beta_i\|_Q^2 - 2\mu_i[(1-\delta_i)\|\gamma_i\|_Q^2] = 0, \quad 1 \leq i \leq r, \quad i \neq n.$$

From this

$$\frac{(1-\delta_n)}{\delta_n}\|\gamma_n\|_Q^2 - \|\beta_n\|_Q^2 - \sum_{\substack{i=1\\ i\neq n}}^{r} \mu_i[\frac{1-\delta_n}{\delta_n}\|\gamma_n\|_Q^2] = 0,$$

and

$$\mu_i = \frac{\|\beta_i\|_Q^2}{\|\gamma_i\|_Q^2}\,\frac{\delta_i}{1-\delta_i}, \quad 1 \leq i \leq r, \quad i \neq n.$$

Substituting μ_i we have:

$$\frac{1-\delta_n}{\delta_n}\|\gamma_n\|_Q^2 - \|\beta_n\|_Q^2 - \sum_{\substack{i=1\\ i\neq n}}^{r} \frac{\|\beta_i\|_Q^2}{\|\gamma_i\|_Q^2}\|\gamma_n\|_Q^2\,\frac{(1-\delta_n)\delta_i}{(1-\delta_i)\delta_n} = 0.$$

Or

$$\frac{1-\delta_n}{\delta_n}\|\gamma_n\|_Q^2 - \sum_{i=1}^{r} \|\beta_i\|_Q^2\,\frac{\|\gamma_n\|_Q^2}{\|\gamma_i\|_Q^2}\,\frac{(1-\delta_n)\delta_i}{(1-\delta_i)\delta_n} = 0.$$

Then

$$\frac{1-\delta_n}{\delta_n}\|\gamma_n\|_Q^2\,[1 - \sum_{i=1}^{r} \frac{\|\beta_i\|_Q^2}{\|\gamma_i\|_Q^2}\,\frac{\delta_i}{1-\delta_i}] = 0.$$

Since Q is a p.d. matrix, $\|\gamma_n\|_Q > 0$ and the last equation is equivalent to

$$\sum_{i=1}^{r} \frac{\|\beta_i\|_Q^2}{\|\gamma_i\|_Q^2}\,\frac{\delta_i}{1-\delta_i} = 1.$$

From the definition of β_i the last formula is equivalent to

$$\sum_{i=1}^{r} \frac{1}{\|\gamma_i\|_T^2} \cdot \frac{\delta_i}{1-\delta_i} = 1,$$

and then the third condition (c) is proved.

It is not immediate that M_cY as defined in the above proposition is the Q-m.e. for the set $C_r = \{\gamma_i\}_{i=1}^r$. The next proposition will show that for any estimator $MY = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y$ there exists a set C_M for which MY is the Q-m.e. As a consequence we obtain that M_cY is indeed the Q-m.e. for C_r.

PROPOSITION 4.6.

Let $MY = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y$ such that $0 < \delta_i < 1$.

(a) Let $C_M = \{\gamma_i\}_{i=1}^r$ a set of T-orthogonal vectors of $\mathbb{R}^p$ which satisfy the following conditions:

(i) $\gamma_i = t_i\beta_i$ for some $t_i \in \mathbb{R}$, $1 \leq i \leq r$.

(ii) $(1-\delta_i)\|\gamma_i\|_Q = (1-\delta_j)\|\gamma_j\|_Q$, $\quad 1 \leq i, j \leq r$.

(iii) $\displaystyle\sum_{i=1}^{r} \frac{\delta_i}{1-\delta_i} \frac{1}{\|\gamma_i\|_T^2} = 1.$

Then MY is the Q-minimax estimator for C_M.

(b) The estimator M_cY defined in Proposition 4.5 is the Q-minimax estimator for $C_r = \{\gamma_i\}_{i=1}^r$.

PROOF. Suppose that MY is not the Q-m.e. for C_M, then from Propositions 4.2 and 4.5, the Q-m.e. M_cY for C_M must have the form

$M_cY = \sum_{i=1}^{q} d_i\beta_i\beta_i'X'V^{-1}Y$, with $q < r$. The indexes of the vectors in $C_M = \{\gamma_i\}_{i=1}^r$ can be taken, without loss of generality, in a way that the indexation in M_cY coincides with the one in MY until $i = q$. From Proposition 4.5 we have that $0 < d_i < 1$, $1 \leq i \leq q$, $(1-d_i)\|\gamma_i\|_Q = (1-d_j)\|\gamma_j\|_Q$, $1 \leq i, j \leq q$ and $\sum_{i=1}^{q} \frac{d_i}{1-d_i} \frac{1}{\|\gamma_i\|_T^2} = 1$. Since we also have (iii) and $r > q$, at least for some $1 \leq n \leq q$ we must have $\frac{\delta_n}{1-\delta_n} < \frac{d_n}{1-d_n}$ and so $\delta_n < d_n$. If M_cY is Q-m.e. for C_M from Proposition 4.4 we must have

$$(1-d_n)\|\gamma_n\|_Q \geq \|\gamma_{q+1}\|_Q ,$$

this implies

$$(1-\delta_n)\|\gamma_n\|_Q > \|\gamma_{q+1}\|_Q,$$

and then for any $0 < \delta < 1$ we have

$$(1-\delta_n)\|\gamma_n\|_Q > (1-\delta)\|\gamma_{q+1}\|_Q .$$

But this contradicts that

$$(1-\delta_n)\|\gamma_n\|_Q = (1-\delta_{q+1})\|\gamma_{q+1}\|_Q , \quad 0 < \delta_{q+1} < 1,$$

which in (ii) we assume it holds. This proves that MY is the Q-m.e. for C_M. Now (b) follows immediately from the fact that if M_cY is the Q-m.e. for C, from Proposition 4.5 it satisfies conditions (i), (ii) and (iii) in (a) and then it must be the Q-m.e. of $\{\gamma_i\}_{i=1}^{r} = C_r = C_M$.

When $0 < \delta_i < 1$, $1 \leq i \leq p$, we know from Theorem 2.4 (i) that $MY = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y$ can be represented as $MY = (T+G)^{-1}X'V^{-1}Y$ where G is a p.d. symmetric matrix and a set of T-eigenvectors of G is $\{\beta_i\}_{i=1}^{p}$. The next proposition gives similar results to Proposition 4.6 for MY represented in this way.

PROPOSITION 4.7.

Let $MY = (T+G)^{-1}X'V^{-1}Y$, G a p.d. symmetric matrix. Suppose that $\{\beta_i\}_{i=1}^p$ is a set of T-orthonormal eigenvectors of G. Then MY is the Q-minimax estimator of the set $C_G = \{\gamma_i\}_{i=1}^p$ defined as follows:

(i) $\gamma_i = t_i\beta_i$, for some $t_i \in \mathbb{R}$, $1 \le i \le p$.

(ii) If $a_i = \gamma_i'(T(T+G)^{-1} - I)Q((T+G)^{-1}T-I)\gamma_i$, then $a_i = a_j$, $1 \le i, j \le p$.

(iii) $\displaystyle\sum_{i=1}^p \frac{1}{\|\gamma_i\|_G^2} = 1.$

PROOF. (i) Follows from the fact that if $MY = \sum_{i=1}^p \delta_i\beta_i\beta_i'X'V^{-1}Y$ the set $\{\beta_i\}_{i=1}^p$ is a set of T-eigenvectors of G and Proposition 4.2. (ii) is only a reformulation of Proposition 4.6 (ii). We will see that (iii) is equivalent to (iii) in Proposition 4.6. Let $G = T(\sum_{i=1}^p \lambda_i\beta_i\beta_i')T$, then from Lemma 2.1(f) and (d) we have $\delta_i = \frac{1}{1+\lambda_i}$ or $\frac{\delta_i}{1-\delta_i} = \frac{1}{\lambda_i}$, $1 \le i \le p$. Therefore Proposition 4.6 (iii) can be written as

$$\sum_{i=1}^p \frac{1}{\|\gamma_i\|_T^2} \cdot \frac{1}{\lambda_i} = 1.$$

Now noticing that $\|\gamma_i\|_T^2 \cdot \lambda_i = \|\gamma_i\|_G^2$ because $\gamma_i = t_i\beta_i$ and β_i is a T-eigenvector of G corresponding to the T-eigenvalue λ_i, (iii) follows.

Since conditions Proposition 4.5 (c) and Proposition 4.7 (iii) are equivalent, the T-cube Z_G for which $(T+G)^{-1}X'V^{-1}Y$ is Q-minimax estimator is totally contained in the G-unit ball, that is in the set $B_G = \{\beta | \beta'G\beta \le 1\}$. The next theorem summarizes the results of some of the previous propositions, gives explicit expression for the shrinkage factors and gives the supremum of the Q-risk of M_CY on C.

THEOREM 4.1.

Let $C = \{\gamma_i\}_{i=1}^m$ a set of T-orthogonal vectors of $\mathbb{R}^p$ and $\beta_i = \frac{\gamma_i}{\|\gamma_i\|_Q}$, $1 \leq i \leq m$. Let $M_cY = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y$. Then

(a) M_cY is the (unique) Q-minimax estimator of C if and only if M_cY satisfies

(i) $0 < \delta_i < 1, \quad 1 \leq i \leq r.$

(ii) $(1-\delta_i)\|\gamma_i\|_Q = (1-\delta_j)\|\gamma_j\|_Q, \quad 1 \leq i, j \leq r.$

(iii) $\sum_{i=1}^{r} \frac{\delta_i}{1-\delta_i} \frac{1}{\|\gamma_i\|_T^2} = 1.$

(iv) $(1-\delta_i)\|\gamma_i\|_Q \geq \|\gamma_k\|_Q, \quad 1 \leq i \leq r, \quad r < k \leq m.$

(b) If M_cY is the Q-minimax estimator of C, its shrinkage factors are given by

$$\delta_n = 1 - \frac{1}{\|\gamma_n\|_Q} \frac{\sum_{i=1}^{r} \frac{\|\gamma_i\|_Q}{\|\gamma_i\|_T^2}}{1 + \sum_{i=1}^{r} \frac{1}{\|\gamma_i\|_T^2}}, \qquad 1 \leq n \leq r.$$

(c) $$\sup_{\gamma\in C} R_Q[M_c,\gamma] = \|\gamma_n\|_Q^2(1-\delta_n)^2[1 + \sum_{i=1}^{r} \frac{\delta_i^2}{(1-\delta_i)^2} \frac{1}{\|\gamma_i\|_T^2}] .$$

PROOF. Proposition 4.6 implies that if M_cY satisfies (a) (i), (ii) and (iii), M_cY is the Q-minimax estimator for $C_r = \{\gamma_i\}_{i=1}^r$. It is also clear that (iv) implies

$$\sup_{\gamma\in C} R_Q[M_c,\gamma] = \sup_{\gamma\in C_r} R_Q[M_c,\gamma] .$$

But this implies that M_c is the Q-minimax estimator for C. Now suppose that M_cY is the Q-m.e. for C, then Proposition 4.5 implies (a)(i), (ii) and (iii). Condition (a)(iv) is implied by Proposition 4.4 (b). The unicity of M_cY is established in Proposition 4.3. This completes the proof of (a). To see (b) from (a)(ii) we have

$$\frac{\delta_i}{1-\delta_i} = \frac{1}{1-\delta_n} \frac{\| \gamma_i \|_Q}{\| \gamma_n \|_Q} - 1, \qquad 1 \leq i, n \leq r.$$

Substituting this expression for each i in (a)(iii) we obtain

$$\sum_{i=1}^{r} \left(\frac{1}{1-\delta_n} \frac{\| \gamma_i \|_Q}{\| \gamma_n \|_Q} - 1\right) \frac{1}{\| \gamma_i \|_T^2} = 1,$$

$$(1-\delta_n) = \sum_{i=1}^{r} \left(\frac{\| \gamma_i \|_Q}{\| \gamma_n \|_Q \| \gamma_i \|_T^2} - \frac{1-\delta_n}{\| \gamma_i \|_T^2} \right),$$

$$(1-\delta_n) \left(1 + \sum_{i=1}^{r} \frac{1}{\| \gamma_i \|_T^2}\right) = \sum_{i=1}^{r} \frac{\| \gamma_i \|_Q}{\| \gamma_n \|_Q \| \gamma_i \|_T^2} .$$

And then

$$\delta_n = 1 - \frac{1}{\| \gamma_n \|_Q} \frac{\sum_{i=1}^{r} \frac{\| \gamma_i \|_Q}{\| \gamma_i \|_T^2}}{1 + \sum_{i=1}^{r} \frac{1}{\| \gamma_i \|_T^2}} .$$

We will now prove (c). From (a)(ii) we obtain

$$\delta_i^2 \| \beta_i \|_Q^2 = (1-\delta_n)^2 \left(\frac{\delta_i}{1-\delta_i}\right)^2 \frac{\| \gamma_n \|_Q^2}{\| \gamma_i \|_T^2}, \qquad 1 \leq i,n \leq r.$$

Therefore

$$\mathrm{Var}_Q[M_c] = \sum_{i=1}^{r} \delta_i^2 \| \beta_i \|_Q^2 = (1-\delta_n)^2 \| \gamma_n \|_Q^2 \sum_{i=1}^{r} \left(\frac{\delta_i}{1-\delta_i}\right)^2 \frac{1}{\| \gamma_i \|_T^2} .$$

From Proposition 4.4, (c), we have

$$\sup_{\gamma \in C} R_Q[M_c, \gamma] = R_Q[M_c, \gamma_n] = (1-\delta_n)^2 \| \gamma_n \|_Q^2 + \mathrm{Var}_Q[M_c] .$$

We then obtain

$$\sup_{\gamma \in C} R_Q[M_c, \gamma] = (1-\delta_n)^2 \| \gamma_n \|_Q^2 \left[1 + \sum_{i=1}^{r} \left(\frac{\delta_i}{1-\delta_i}\right)^2 \frac{1}{\| \gamma_i \|_T^2}\right] .$$

The theorem provides us with the following way of calculating the Q-minimax estimator M_cY of an arbitrary set of T-orthogonal vectors $C = \{\gamma_i\}_{i=1}^m$. Let first order C in such a way that $\| \gamma_1 \|_Q \geq \| \gamma_2 \|_Q \geq \ldots \geq \| \gamma_m \|_Q$ and calculate the Q-m.e. M_1Y for $C_1 = \{\gamma_i\}_{i=1}^1$. Then if

$$B_Q[M_1, \gamma_1] \geq \| \gamma_i \|_Q^2 , \qquad 1 < i \leq m, \tag{4.1.11}$$

Theorem 4.1, (a) (iv) says that M_1Y is in fact M_cY. If (4.1.11) does not hold we calculate the Q-m.e. M_2Y for $C_2 = \{\gamma_i\}_{i=1}^2$ using Theorem 4.1 (b). If

$$B_Q[M_2, \gamma_2] \geq \| \gamma_i \|_Q^2 , \qquad 2 < i \leq m, \tag{4.1.12}$$

then from Theorem 4.1, (a) (iv), we know that M_2Y is M_cY. If (4.1.12) does not hold we do the process until the step $r \leq m$ where

$$B_Q[M_r, \gamma_r] \geq \| \gamma_i \|_Q^2, \qquad r < i \leq m, \tag{4.1.13}$$

holds. From Theorem 4.1, (a) (iv), we know that M_rY is M_cY. The only remaining point to clear up in this method is that if at the step $n < m$ we have

$$B_Q[M_n, \gamma_n] < \| \gamma_{n+1} \|_Q^2, \qquad (4.1.14)$$

then the Q-m.e. $M_{n+1}Y$ of C_{n+1} must have all its n+1 shrinkage factors different from zero. Let us suppose the contrary and let W a proper subset of {1 ..., n+1} such that $M_W Y = \sum_{i \in W} \delta_i \beta_i \beta_i' X' V^{-1} Y$ is the Q.m.e. of $C_{n+1}, \delta_i \neq 0$, $i \in W$. From Proposition 4.6(b), $M_W Y$ is the Q-m.e. for $C_W = \{\gamma_i | i \in W\}$. Let k the smallest number which satisfies $1 \leq k \leq n+1$ and $k \notin W$. We will see that W cannot be contained in {1,... k-1}. If we suppose the contrary from Proposition 4.6, (b) and the condition on k $M_W Y$ is the Q-m.e. for C_{k-1}. Also because $M_W Y$ is the Q-m.e. for C_{n+1}, from Proposition 4.4, (b) we have

$$B_Q[M_W, \gamma_{k-1}] \geq \| \gamma_i \|_Q^2, \qquad k \leq i \leq n+1,$$

and then from Theorem 4.1, (a) (iv) $M_W Y$ is also Q-m.e. for C_n, by unicity we then have $M_W Y = M_n Y$; but then from the last equality we have

$$B_Q[M_n, \gamma_{k-1}] \geq \| \gamma_{n+1} \|_Q^2 .$$

This last inequality contradicts (4.1.14); then W is not contained in {1,...,k-1}. From this, the condition on k and the definition of $M_W Y$ there exist $k < j \leq n+1$ such that $\delta_j > 0$. From Proposition 4.4, (b) we have

$$\| \gamma_j \|_Q^2 > B_Q[M_W, \gamma_j] \geq \| \gamma_k \|_Q^2 .$$

This contradicts the assumption on the ordering of the γ_i's made at the beginning of the method. This proves that all the n+1 shrinkage factors of $M_{n+1}Y$ will be different from zero and then Theorem 4.1,(b) can be applied to calculate them.

When $\|\gamma_i\|_Q = \|\gamma_j\|_Q$, $1 \leq i, j \leq m$, some important simplifications can be done. This particular case will be studied in the next corollary.

COROLLARY 4.1.1.

Let $C = \{\gamma_i\}_{i=1}^m$ a set of T-orthogonal vectors of $\mathbb{R}^p$ such that

$$\|\gamma_i\|_Q^2 = \ell \quad , \qquad 1 \leq i \leq m.$$

Let $\beta_i = \dfrac{\gamma_i}{\|\gamma_i\|_T}$, $1 \leq i \leq m$ and $a = \displaystyle\sum_{i=1}^m \frac{1}{\|\gamma_i\|_T^2}$. Then

(a) $M_c Y = \dfrac{1}{1+a} \hat{\beta}^S$, where $S = \dot{S}\text{pan}\,(C)$.

(b) $\displaystyle\sup_{\gamma \in C} R_Q[M_c, \gamma] = \ell \frac{a}{1+a}$.

PROOF. From (a)(ii) in the theorem we have $\delta = \delta_i$, $1 \leq i \leq r$. Since

$$\sum_{i=1}^r \frac{\delta_i}{1-\delta_i} \frac{1}{\|\gamma_i\|_T^2} = \frac{\delta}{1-\delta} \sum_{i=1}^r \frac{1}{\|\gamma_i\|_T^2} ,$$

can be made equal to 1, by appropriate choice of δ for an arbitrary r, we have that $r = m$ and

$$\frac{\delta}{1-\delta} \sum_{i=1}^m \frac{1}{\|\gamma_i\|_T^2} = 1 ,$$

or

$$\delta = \frac{1}{1+a} .$$

Then $M_c Y = \delta \displaystyle\sum_{i=1}^m \beta_i \beta_i' X' V^{-1} Y$. Since $\beta_i = \dfrac{\gamma_i}{\|\gamma_i\|_T}$, $\text{Span}\{\beta_i\}_{i=1}^m =$ Span (C) = S and from Theorem 2.2, (i), we know that $\displaystyle\sum_{i=1}^m \beta_i \beta_i' X' V^{-1} Y = \hat{\beta}^S$.

This proves (a). A simple calculation shows that (b) follows when δ_n and δ_i are substituted in Theorem 4.1, (c) by $\frac{1}{1+a}$.

We will now study some particular examples where the previous results are applied.

EXAMPLE 4.1.

Let $Q = I$, $\{\beta_i\}_{i=1}^p$ a set of T-eigenvectors of I, $\gamma_i = t_i \beta_i$, $\| \gamma_i \|_I^2 = t$, $1 \le i \le p$ and $C = \{\gamma_i\}_{i=1}^p$. Let first notice that $\| \gamma_i \|_T^2 = \lambda_i \| \gamma_i \|_I^2 = \lambda_i t$, where the λ_i's are the eigenvalues of T. Then from the corollary we have

$$a = \sum_{i=1}^{p} \frac{1}{\| \gamma_i \|_T^2} = \sum_{i=1}^{p} \frac{1}{t.\lambda_i} = \frac{1}{t} . \operatorname{Tr}[T^{-1}].$$

Then

$$\delta = \frac{t}{t + \operatorname{Tr}[T^{-1}]} .$$

And

$$M_c Y = \frac{t}{t + \operatorname{Tr}[T^{-1}]} \hat{\beta} .$$

And

$$\sup_{\gamma \in C} R_Q[M_c, \gamma] = t \frac{a}{1+a} = \frac{t.\operatorname{Tr}[T^{-1}]}{t + \operatorname{Tr}[T^{-1}]} .$$

EXAMPLE 4.2.

Let C as in Example 4.1, but $Q = \sigma^2 T$ (the Q-risk corresponding to $Q = \sigma^2 T$ is usually called the prediction mean square error, see Brown et al. (1978)). In this example if T is very ill conditioned it can happen that $r < p$ and then the method described at the end of Theorem 4.1 should be applied to find $M_c Y$.

EXAMPLE 4.3.

Let $Q = \sigma^2 T$, $\{\beta_i\}_{i=1}^p$ a set of T-eigenvectors of I and $\gamma_i = t^{\frac{1}{2}}\beta_i$, $1 \leq i \leq p$. Let $C = \{\gamma_i\}_{i=1}^p$. As in Example 4.1 we must have $\delta = \delta_i$, $1 \leq i \leq p$. Since $\|\gamma_i\|_T^2 = t\|\beta_i\|_T^2 = t$ we have

$$a = \sum_{i=1}^{p} \frac{1}{\|\gamma_i\|_T^2} = \frac{p}{t} .$$

From this

$$\delta = \frac{t}{t+p} ,$$

$$M_c Y = \frac{t}{t+p} \hat{\beta} ,$$

and since $\|\gamma\|_Q^2 = \sigma^2\|\gamma\|_T^2 = \sigma^2 t$ for $\gamma \in C$

$$\sup_{\gamma \in C} R_Q[M_c, \gamma] = \sigma^2 \frac{tp}{t+p} .$$

The next theorem extends the results of Theorem 4.1 to limits of T-orthogonal sets.

THEOREM 4.2.

Let $C = \{\gamma_i\}_{i=1}^r \cup (\bigcup_{i=r+1}^{m} S_i) = C_1 \cup C_2$ such that C_1 is a set of T-orthogonal vectors of $\mathbb{R}^p$, S_i is a subspace of dimension 1 for $r < i \leq m$, $S_i \perp_T S_j$, $r < i, j \leq m$ and $C_1 \perp_T C_2$. Let $\beta_i = \frac{\gamma_i}{\|\gamma_i\|_T}$, $1 \leq i \leq r$ and $\beta_i \in S_i$, $\|\beta_i\|_T = 1$, $r < i \leq m$.

Let $S = \mathrm{Span}(C_2)$ and $\hat{\beta}^s$ the S-BLUE. Then

(a) There exists a unique Q-minimax estimator $M_c Y$ for C given by

$$M_c Y = M_1 Y + \hat{\beta}^s,$$

where $M_1 Y$ is the Q-minimax estimator for C_1.

(b) (i) $\sup_{\gamma\in C} R_Q[M_c,\gamma] = \sup_{\gamma\in C_1} R_Q[M_c,\gamma]$.

(ii) $\sup_{\gamma\in C} R_Q[M_c,\gamma] = \sup_{\gamma\in C_1} R_Q[M_1,\gamma] + \mathrm{Var}_Q[\hat{\beta}^s]$.

PROOF. Proposition 4.2 can easily be extended to sets C of the form given in this theorem and then there is no loss of generality in considering M_cY as $M_cY = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y + \sum_{i=r+1}^{m} \delta_i\beta_i\beta_i'X'V^{-1}Y = M_1Y+M_2Y$. Now if $\delta_i \neq 1$ for some $r < i \leq m$ we have that $\sup_{\gamma\in C} R_Q[M_c,\gamma] = \infty$ but $\sup_{\gamma\in C} R_Q[\hat{\beta},\gamma] = \mathrm{Tr}[QT^{-1}] < \infty$ and therefore M_cY is not the Q-m.e. for C. We then have $\delta_i=1$, $r<i\leq m$. Theorem 2.2, (i) implies that $M_2Y = \beta^s$. Then for any $\gamma_i \in C_1$ we have

$$R_Q[M_c,\gamma_i] = (1-\delta_i)^2\|\gamma_i\|_Q^2 + \mathrm{Var}_Q[M_c] \geq \mathrm{Var}_Q[M_c] = \sup_{\gamma\in C_2} R_Q[M_c,\gamma] \quad ,$$

and then

$$\sup_{\gamma\in C} R_Q[M_c,\gamma] = \sup_{\gamma\in C_1} R_Q[M_c,\gamma] \quad .$$

This proves (b) (i). Since

$$R_Q[M_c,\gamma] = B_Q[M_c,\gamma] + \mathrm{Var}_Q[M_1] + \mathrm{Var}_Q[\hat{\beta}^s],$$

and $B_Q[M_c,\gamma] = 0$, for $\gamma \in C_2$, we have that

$$\sup_{\gamma\in C} R_Q[M_c,\gamma] = \sup_{\gamma\in C_1} R_Q[M_1,\gamma] + \mathrm{Var}_Q[\hat{\beta}^s] \quad .$$

This proves (b) (ii). This also implies that M_cY is Q-m.e. over C if and only if M_1Y is the Q-m.e. over C_1. This proves (a) and the theorem is proved. As a consequence we have

COROLLARY 4.2.1.

Let $C = \{\gamma_i\}_{i=1}^{r} \cup S = C_1 \cup S$, such that C_1 is a set of T-orthogonal vectors of $\mathbb{R}^p$, S is a subspace of $\mathbb{R}^p$ and $S \perp_T C_1$. Then the

Q-minimax estimator M_cY is given by

$$M_cY = M_1Y + \hat{\beta}^s,$$

where M_1Y is the Q-minimax estimator for C_1.

PROOF. Let $C_2 = \bigcup_{i=r+1}^{m} S_i$ as defined in the theorem, then $\mathrm{Span}(C_2) = S$ and M_cY is the Q-m.e. for $C_1 \cup C_2$. The result of the corollary follows from the fact that

$$\sup_{\gamma \in C_1 \cup C_2} R_Q[M_c,\gamma] = \sup_{\gamma \in C} R_Q[M_c,\gamma] \quad .$$

From Proposition 4.6 and the last theorem it can be seen that to any estimator of the form $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y, 0 \leq \delta_i \leq 1, 1 \leq i \leq p$ and $\{\beta_i\}_{i=1}^{p}$ a set of T-orthonormal vectors of $\mathbb{R}^p$; a set C_M is associated for which MY is the unique Q-minimax estimator. We will use this fact to give an important property of the set of ALE's in the next proposition.

PROPOSITION 4.8.

Let Q a p.d. symmetric matrix. Let $MY \in L$ be of the form $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ where $0 \leq \delta_i \leq 1$ for $1 \leq i \leq p$ and $\{\beta_i\}_{i=1}^{p}$ is a T-orthonormal set of vectors of $\mathbb{R}^p$. Then MY is an ALE under the risk defined by Q.

PROOF. It is enough to prove that there does not exist an $M_0Y \in L$ and a $\gamma_0 \in \mathbb{R}^p$ such that

$$R_Q[M_0,\gamma] \leq R_Q[M,\gamma] \quad , \qquad \forall \gamma \in \mathbb{R}^p ,$$

and

$$R_Q[M_0,\gamma_0] < R_Q[M,\gamma_0].$$

Let C_M the set of Proposition 4.6 or Theorem 4.2 for which MY is the unique Q-m.e. From the unicity of MY, the first inequality implies that $M_0Y = MY$ but then the second inequality is not true. This proves the proposition.

If MY is the unique Q-m.e. for a set C and $\sup_{\gamma \in C} R_Q[M,\gamma] = K_c$, then MY is the unique Q-m.e. for any set C_1 which contains C and such that $\sup_{\gamma \in C_1} R_Q[M,\gamma] = K_c$. In this way using Theorem 4.2 and Proposition 4.6 for each estimator MY of the form given in the last proposition a unique ellipsoid $\{\gamma | R_Q[M,\gamma] \leq \sup_{\gamma \in C_M} R_Q[M,\gamma]\}$ can be associated. This suggests the next definition and Proposition 4.9.

DEFINITION 4.3.

The Q-bias matrix B_M^Q of the estimator MY is

$$B_M^Q = (X'M'-I)Q(MX-I).$$

The Q-bias ellipsoid of radius R, $B_M^Q(R)$ of an estimator MY will be

$$B_M^Q(R) = \{\gamma | \gamma' B_M^Q \gamma \leq R^2\}.$$

Using this definition we have that

$$B_Q[M,\gamma] = \gamma' B_M^Q \gamma .$$

PROPOSITION 4.9.

Let MY be an ALE. Then there exist at least one value R for which MY is the unique Q-minimax estimator for $B_M^Q(R)$.

PROOF. From Theorem 2.2 if MY is an ALE it has the form given in Proposition 4.8. Then from Theorem 4.2 and Proposition 4.6 there exists a set C_M for which MY is the unique Q-m.e. Let $R = \sup_{\gamma \in C_M} B_Q[M,\gamma]$ then the proposition follows from the fact that

$B_M^Q(R) = \{\gamma | R_Q[M,\gamma] \leq \sup_{\gamma \in C_M} R_Q[M,\gamma]\}$ and that $B_M^Q(R)$ contains C_M.

Proposition 3.10 essentially says that if we are looking for Q-minimax estimators we need only to consider ALE's. Prop. 4.6 and Theorem 4.2 are a "reverse" to this statement. The next theorem, which is a new characterization for ALE's, "summarizes" those results.

THEOREM 4.3.

Let $MY \in L$. The following statements are equivalent.

(i) There exists a subset C of $\mathbb{R}^p$ for which MY is the unique Q - minimax estimator .

(ii) MY is an ALE.

PROOF. The theorem is a direct consequence of Proposition 3.10, Proposition 4.6 , Theorem 4.2 and Theorem 2.3.

4.2. Q-MINIMAX ESTIMATORS OVER ELLIPSOIDS CENTRED AT THE ORIGIN

The problem of finding Q-minimax estimators for ellipsoids centred at the origin is not new. Lauter in 1975 gave a representation for those estimators. The approach of this thesis allows clarification of a number of points arising from Lauter's work:

(i) The expression given by Lauter is not very tractable when an actual Q-minimax estimator MY has to be calculated.

(ii) Given an estimator MY it would be convenient to be able to discover on what ellipsoids this estimator is Q-minimax (we know from Proposition 4.9 that if MY is an ALE there exist at least one). This "reverse" problem has not been treated by Lauter.

Whenever the word ellipsoid will be used in the section it will mean ellipsoid centred at the origin.

The section will be organized as follows:

Proposition 4.10 shows that a linear estimator cannot be a Q-minimax estimator over two ellipsoids with constant Q-risk on their border. The ellipsoid over which an ALE is Q-minimax estimator and has constant Q-risk on its border is characterized in Theorem 4.4 and Theorem 4.5. Corollary 4.4.1 is a characterization of the ellipsoid when the ALE is represented in its usual form; it is useful because in many situations ALE's are given in such representation. Theorem 4.6 characterizes the ALE's which are Q-minimax estimators over only one ellipsoid. Lemma 4.1 plays an important part to obtain this result and Theorem 4.7. Corollary 4.6.1 translates Theorem 4.6 in terms of the usual representation of ALE's. Then some examples are studied where those results are applied, those examples include ridge estimators, generalized ridge estimators and estimators of the form $t\hat{\beta}$. A small digression is made just to show possible connections of this work with the "Empirical Bayes Rules". Finally Theorem 4.7 gives an explicit procedure to calculate Q-minimax estimators over a wide class of ellipsoids. The section is closed with some examples where Theorem 4.7 is applied. The next proposition is just the completion of Proposition 4.9 and will play an essential part in obtaining the main results of the section.

PROPOSITION 4.10.

Let $MY = AX'V^{-1}Y$ an ALE. Then there exists a unique value R, which depends on M and Q, such that MY is the (unique) Q-minimax estimator for $B_M^Q(R)$.

PROOF. From Proposition 4.9 we only need to prove the unicity of R. Suppose that there exist $R_0 \neq R$ such that MY is also Q-m.e. for $B_M^Q(R_0)$. If $A = \sum_{i=1}^{p} \delta_i \beta_i \beta_i'$ let $A_\mu = \sum_{i=1}^{p} d_i \beta_i \beta_i'$ with $\mu(1-\delta_i) = (1-d_i)$ and $M_\mu = A_\mu X'V^{-1}$. We then have

$$\sup_{\gamma \in B_M^Q(\lambda)} B_Q[M_\mu,\gamma] = \mu^2 \sup_{\gamma \in B_M^Q(\lambda)} B_Q[M,\gamma] = \mu^2 . \lambda^2 ,$$

and

$$\mathrm{Var}_Q[M_\mu] = \sum_{i=1}^{p} (1 - \mu(1-\delta_i))^2 \, \| \beta_i \|_Q^2 .$$

Therefore

$$\sup_{\gamma \in B_M^Q(\lambda)} R_Q[M_\mu,\gamma] = \mu^2\lambda^2 + \sum_{i=1}^{p} (1 - \mu(1-\delta_i))^2 \| \beta_i \|_Q^2 = f(\mu,\lambda).$$

Then a necessary condition for $M_\mu Y$ to be Q-m.e. over $B_M^Q(\lambda)$ is

$$\frac{\partial}{\partial \mu} f(\mu,\lambda) = 0.$$

This is equivalent to

$$\mu(\lambda^2 + \sum_{i=1}^{p} (1-\delta_i)^2 \|\beta_i\|_Q^2) = \sum_{i=1}^{p} (1-\delta_i)\|\beta_i\|_Q^2.$$

Then if MY is Q-m.e. for $B_M^Q(R)$ and $B_M^Q(R_0)$, since MY is $M_\mu Y$ when $\mu = 1$, we must have that μ satisfies simultaneously

$$\mu(R^2 + \sum_{i=1}^{p} (1-\delta_i)^2 \|\beta_i\|_Q^2) = \sum_{i=1}^{p} (1-\delta_i)\|\beta_i\|_Q^2 ,$$

$$\mu(R_0^2 + \sum_{i=1}^{p} (1-\delta_i)^2 \|\beta_i\|_Q^2) = \sum_{i=1}^{p} (1-\delta_i)\|\beta_i\|_Q^2 ,$$

and $\mu = 1$. But this is impossible if $R \neq R_0$. This proves the proposition. The next theorem characterizes this unique ellipsoid for a wide class of ALE's and gives the Q-risk on its boundary.

THEOREM 4.4.

Let $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ an ALE with $0 < \delta_i < 1$, $1 \leq i \leq p$.

then

(a) MY is the unique Q-minimax estimator over the ellipsoid

$$C_M^Q = B_M^Q \left(\left(\sum_{i=1}^{p} (1-\delta_i)\delta_i \|\beta_i\|_Q^2 \right)^{\frac{1}{2}} \right).$$

(b) The value $R_Q[M]$ of the Q-risk of MY on the boundary of C_M^Q is

$$R_Q[M] = \sum_{i=1}^{p} \delta_i \|\beta_i\|_Q^2 .$$

PROOF. The Q-bias matrix for MY is

$$B_M^Q = T\left(\sum_{1 \leq i,k \leq p} \beta_i' Q \beta_k (1-\delta_i)(1-\delta_k) \beta_i \beta_k' \right) T.$$

Let $\gamma_i = \dfrac{R}{(1-\delta_i)\|\beta_i\|_Q} \cdot \beta_i$. Then it is easy to check that

$$B_Q[M,\gamma_i] = \gamma_i{}'B_M^Q\,\gamma_i = R^2.$$

Now if the γ_i's satisfy the additional condition

$$\sum_{i=1}^{p} \frac{\delta_i}{1-\delta_i}\,\frac{1}{\|\gamma_i\|_T^{\,2}} = 1,$$

from Proposition 4.6, (a), we know that MY will be the Q-m.e. for $\{\gamma_i\}_{i=1}^{p}$. Since from Proposition 4.10 there is only one R such that MY is the Q-m.e. for $B_M^Q(R)$, R must satisfy

$$\sum_{i=1}^{p} \frac{\delta_i}{1-\delta_i}\,\frac{(1-\delta_i)^2}{R^2}\,\|\beta_i\|_Q^{\,2} = 1.$$

From this (a) is readily obtained. From the above equality we have

$$R^2 = \sum_{i=1}^{p} (1-\delta_i)\delta_i\|\beta_i\|_Q^{\,2}\ .$$

Then for any γ belonging to the boundary of C_M^Q we have

$$B_Q[M,\gamma] = \sum_{i=1}^{p} (1-\delta_i)\delta_i\|\beta_i\|_Q^{\,2}$$

Since $\mathrm{Var}_Q[M] = \sum_{i=1}^{p} \delta_i^2\|\beta_i\|_Q^{\,2}$, we have for any γ in the boundary of C_M^Q

$$R_Q[M,\gamma] = \sum_{i=1}^{p} (1-\delta_i)\delta_i^2\|\beta_i\|_Q^{\,2} + \sum_{i=1}^{p} \delta_i^{\,2}\|\beta_i\|_Q^{\,2} = \sum_{i=1}^{p} \delta_i\|\beta_i\|_Q^{\,2}\ .$$

This proves (b) and the theorem is proved.

The next corollary translates the result of the theorem to the "non-shrinkage" representation of MY.

COROLLARY 4.4.1.

Let G a p.d. symmetric p×p matrix. Let $MY = (T+G)^{-1}X'V^{-1}Y$. Then

(a) MY is the unique Q-minimax estimator for

$$C_M^Q = B_M^Q\left((\mathrm{Tr}[B_M^Q G^{-1}])^{\frac{1}{2}}\right) .$$

(b) The Q-risk $R_Q[M]$ at any point of the boundary of C_M^Q is

$$R_Q[M] = \mathrm{Tr}[(T+G)^{-1}Q].$$

PROOF. From Theorem 2.4 if $MY = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y$ we must have

$G = T\left(\sum_{i=1}^{p} \frac{1-\delta_i}{\delta_i}\beta_i\beta_i'\right)T$. Then from Lemma 2.1,(d), $G^{-1} = \sum_{i=1}^{p} \frac{\delta_i}{1-\delta_i}\beta_i\beta_i'$.

Therefore

$$B_M^Q G^{-1} = T\left(\sum_{1\le i,k\le p} \beta_i'Q\beta_k(1-\delta_i)(1-\delta_k)\beta_i\beta_k'\right)T\left(\sum_{i=1}^{p} \frac{\delta_i}{1-\delta_i}\beta_i\beta_i'\right).$$

Using the T-orthonormality of $\{\beta_i\}_{i=1}^{p}$ we have

$$B_M^Q G^{-1} = T\left(\sum_{i=1}^{p} (1-\delta_i)\delta_i\|\beta_i\|_Q^2\,\beta_i\beta_i'\right) .$$

Therefore

$$\mathrm{Tr}[B_M^Q G^{-1}] = \sum_{i=1}^{p} (1-\delta_i)\delta_i\|\beta_i\|_Q^2 .$$

Then (a) is implied from Theorem 4.4, (a). To see (b) notice first that

$$(T+G)^{-1} = \sum_{i=1}^{p} \delta_i\beta_i\beta_i' .$$

And then

$$\mathrm{Tr}[(T+G)^{-1}Q] = \mathrm{Tr}\left[\sum_{i=1}^{p} \delta_i\beta_i\beta_i'Q\right] = \sum_{i=1}^{p} \delta_i\|\beta_i\|_Q^2 = R_Q[M].$$

This proves (b).

The next theorem generalizes Theorem 4.4 to all the ALE's.

THEOREM 4.5.

Let $MY = \sum_{i=1}^{r} \delta_i\beta_i\beta_i'X'V^{-1}Y + \sum_{i=r+1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y = M_1Y+\hat{\beta}^s$ an ALE.

with $0 \le \delta_i < 1$, $1 \le i \le r$ (and $\delta_i = 1$ for $r<i \le p$). Then

(a) MY is the unique Q-minimax estimator for the ellipsoid

$$C_M^Q = B_M^Q((\sum_{i=1}^{p} (1-\delta_i)\delta_i \| \beta_i \|_Q^2)^{\frac{1}{2}}) .$$

(b) The value $R_Q[M]$ of the Q-risk of MY on the boundary of C_M^Q is

$$R_Q[M] = \sum_{i=1}^{p} \delta_i \| \beta_i \|_Q^2 .$$

PROOF. The argument of Theorem 4.4 can be followed through to see that if $R^2 = (\sum_{i=1}^{r} (1-\delta_i)\delta_i \| \beta_i \|_Q^2)$ and $\gamma_i = \frac{R}{(1-\delta_i)\| \beta_i \|_Q} \cdot \beta_i$

$1 \le i \le r$, the estimator MY is the unique Q-m.e for $\{\gamma_i\}_{i=1}^{r}$.

If $C = \{\gamma | R_Q[M,\gamma] \le R_Q[M,\gamma_1]\}$, we also have that MY is the unique Q-m.e. for C. Since

$$B_M^Q((\sum_{i=1}^{r} (1-\delta_i)\delta_i \| \beta_i \|_Q^2)^{\frac{1}{2}}) = C ,$$

(a) follows from the fact that $\sum_{i=1}^{r} (1-\delta_i)\delta_i \| \beta_i \|_Q^2 = \sum_{i=1}^{p} (1-\delta_i)\delta_i \| \beta_i \|_Q^2$.

To prove (b) we have that

$$R_Q[M] = R_Q[M_1] + \sum_{i=r+1}^{p} \| \beta_i \|_Q^2 .$$

Following similar steps as in Theorem 4.4 (b), it can be seen that

$$R_Q[M_1] = \sum_{i=1}^{r} \delta_i \| \beta_i \|_Q^2 .$$

(b) follows from the fact that $\delta_i = 1$, $r < i \le p$.

So far we have proved that for any ALE MY there is a unique ellipsoid C_M^Q for which MY is Q-m.e. and $R_Q[M,\gamma]$ is constant for γ on the boundary of C_M^Q. The question now arises if there is no other ellipsoid C where MY could be Q-m.e. (with, of course, non constant Q-risk of MY on the boundary of C). To treat this question we will first prove a useful lemma. Given a n.n.d. p×p symmetric matrix π, the unit ball of π will be denoted by B_π, that is $B_\pi = \{\gamma | \gamma'\pi\gamma \le 1\}$. We have proved in Theorems 4.4 and 4.5 that if $\pi = \frac{B_M^Q}{R^2}$ then MY is the unique Q-m.e. for B_π. We will see in the next lemma that for any π there exist a unique estimator MY of the form $M\,Y = AX'V^{-1}Y$, with A symmetric, such that $\pi = \frac{B_M^Q}{R^2}$.

LEMMA 4.1.

Let π a n.n.d. symmetric p×p matrix. Then there exists a unique symmetric matrix $A = \sum_{i=1}^{p} \delta_i \beta_i \beta_i'$ ($\{\beta_i\}_{i=1}^{p}$ is a T-orthonormal set of $\mathbf{R}^p$) such that $\delta_i \le 1$, $1 \le i \le p$ and

$$\pi = \frac{(I-TA)Q(I-AT)}{\sum_{i=1}^{p} (1-\delta_i)\delta_i \|\beta_i\|_Q^2} = \frac{B_M^Q}{R^2} \tag{4.2.1}$$

(Here $M = AX'V^{-1}$). Let H the unique n.n.d. symmetric matrix which satisfies $T^{-1}\pi T^{-1} = HQH$. Then we have

$$R = \frac{Tr[HQ]}{1 + Tr[\pi T^{-1}]} , \tag{4.2.2}$$

and

$$A = T^{-1} - RH \tag{4.2.3}$$

PROOF. We have

$$(I-TA)Q(I-AT) = T(T^{-1}-A)Q(T^{-1}-A)T$$

Then if (4.2.1) holds

$$R^2T^{-1}\pi T^{-1} = (T^{-1}-A)Q(T^{-1}-A). \qquad (4.2.4)$$

Given a matrix H, we will call a Q-representation of H a set $\{(\lambda_i, q_i)\}_{i=1}^p$, such that $H = \sum_{i=1}^p \lambda_i q_i q_i'$ and the set $\{q_i\}_{i=1}^p$ is a set of Q-orthonormal vectors. From Lemma 2.1 at least one of such representations exist if H is a symmetric matrix. If H is also n.n.d. then $\lambda_i \geq 0$, $1 \leq i \leq p$. It is evident that H = K if and only if H and K have the same set of Q-representations. If $T^{-1}\pi T^{-1} = HQH$ and π and H are n.n.d. symmetric matrices then the equality

$$\sum_{i=1}^p \lambda_i^2 q_i q_i' = (\sum_{i=1}^p \lambda_i q_i q_i')Q(\sum_{i=1}^p \lambda_i q_i q_i')$$

establishes in the obvious way a bijection between the Q-representations of $T^{-1}\pi T^{-1}$ and those of H. Then if $T^{-1}\pi T^{-1} = KQK$ and K is a n.n.d. symmetric matrix the set of Q-representations of K is the same as the one of H, hence K = H. Then there is a unique n.n.d. symmetric matrix H such that

$$T^{-1}\pi T^{-1} = HQH \quad .$$

Since $\delta_i \leq 1$, $1 \leq i \leq p$, we have $1 - \delta_i \geq 0$, $1 \leq i \leq p$ and therefore

$$T^{-1} - A = \sum_{i=1}^p (1 - \delta_i)\beta_i\beta_i'$$

is a n.n.d. matrix. From this, the uniqueness of H and (4.2.4) we have $RH = T^{-1}-A$. This proves (4.2.3). Now

$$R^2 = \sum_{i=1}^{p} (1-\delta_i)\delta_i \| \beta_i \|_Q^2 = \mathrm{Tr}[(\sum_{i=1}^{p} (1-\delta_i)\beta_i\beta_i')T(\sum_{i=1}^{p} \delta_i\beta_i\beta_i')Q],$$

$$= \mathrm{Tr}[(T^{-1}-A)TAQ] = \mathrm{Tr}[RHT(T^{-1}-RH)Q],$$

$$= R\,\mathrm{Tr}[HQ] - R^2\,\mathrm{Tr}[HTHQ],$$

$$= R\,\mathrm{Tr}[HQ] - R^2\mathrm{Tr}[\pi T^{-1}] \quad .$$

From this we obtain

$$R = \frac{\mathrm{Tr}[HQ]}{1 + \mathrm{Tr}[\pi T^{-1}]} \quad .$$

This is (4.2.2) and the lemma is proved. It should be noticed that we do not have in general $0 < \delta_i$. In fact for (4.2.3) to hold, δ_i may sometimes have to be negative for certain $1 \leq i \leq p$. Now we will characterize the ALE's which are Q-m.e. for only one ellipsoid.

THEOREM 4.6.

Let $MY = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y$. Then the following statements are equivalent:

(a) There is only one ellipsoid (centred at the origin) for which MY is the Q-minimax estimator.

(b) $0 < \delta_i \leq 1, \quad 1 \leq i \leq p$.

PROOF. Let $\pi = \frac{B_M^Q}{R}$ and D a n.n.d. symmetric p×p matrix. Let B_D, ∂B_π and ∂B_D denote the unit ball of D, and the boundaries of B_π and B_D respectively. Suppose that MY is Q-m.e. for B_D. We have three possibilities: $B_D \not\subset B_\pi$; $B_D \subset B_\pi$ and $\partial B_D \cap \partial B_\pi = \phi$ and finally $B_D \subset B_\pi$ and $\partial B_D \cap \partial B_\pi \neq \phi$. Let $a = \sup_{\gamma \in B_D} R_Q[M,\gamma]$ and C the ellipsoid defined as $C = \{\gamma | R_Q[M,\gamma] \leq a\}$. Then if $B_D \not\subset B_\pi$, we will have that MY is Q-m.e. on B_π and C, the Q-risk of MY on ∂B_π and on ∂C is constant and $B_\pi \neq C$. This contradicts Proposition 4.10. Similarly, if $B_D \subset B_\pi$ and $\partial B_D \cap \partial B_\pi = \phi$ we reach a contradiction with Proposition 4.10. The last possibility we are left with is $\partial B_D \cap \partial B_\pi \neq \phi$ and $B_D \subset B_\pi$. Let $\{\pi_n\}_{n=1}^{\infty}$ such that π_n is a n.n.d. symmetric matrix for $n \geq 1$, π_n tends to π when n tends to infinity and if B_n is the unit ball of π_n we have $B_D \subset B_n \subset B_\pi$, for $n \geq 1$. From Lemma 4.1, for each π_n there exist a unique matrix A_n (defined by the Lemma) such that $A_n = \sum_{i=1}^{p} \delta_i^n \beta_i^n \beta_i^{n\prime}$ and $\delta_i^n \leq 1$, $1 \leq i \leq p$. Now since π_n tends to π, from the uniqueness of A_n and A we have that A_n tends to A. Since $0 < \delta_i$ for $1 \leq i \leq p$, there exist k such that if $n > k$, $\delta_i^n > 0$ for $1 \leq i \leq p$; but then the estimator $M_n Y = A_n X'V^{-1}Y$ is an ALE (for $n > k$) and from Theorems 4.4 and 4.5 $M_n Y$ is the Q-m.e. for B_n. Also

$$\sup_{\gamma \in B_D} R_Q[M_n, \gamma] \leq \sup_{\gamma \in B_n} R_Q[M_n, \gamma]\text{, because } B_D \subset B_n.$$

$$\sup_{\gamma \in B_n} R_Q[M_n, \gamma] < \sup_{\gamma \in B_n} R_Q[M, \gamma]\text{, because } M_n Y \text{ is the unique Q-m.e. for } B_n.$$

$$\sup_{\gamma \in B_n} R_Q[M, \gamma] = \sup_{\gamma \in B_D} R_Q[M, \gamma]\text{, because } \partial B_D \cap \partial B_\pi \neq \phi,$$

$B_D \subset B_n \subset B_\pi$ and the Q-risk of MY on ∂B_π is constant. Joining those inequalities we obtain

$$\sup_{\gamma \in B_D} R_Q[M_n, \gamma] < \sup_{\gamma \in B_D} R_Q[M, \gamma] \quad .$$

Thus MY is not the Q-m.e. for B_D. This proves (b) implies (a). To see the other implication suppose $MY = \sum_{i=1}^{r} \delta_i \beta_i \beta_i' X'V^{-1}Y$ be an ALE with $0 < \delta_i$, $1 \leq i \leq r$ and $r < p$. Let $S = \text{Span}\{\beta_i\}_{i=1}^{r}$. Then following the arguments given in Theorems 4.4 and 4.5 it can be seen that MY is the Q-m.e. for C_M^Q and for $C_M^Q \cap S$. Then MY is the Q-m.e. for any ellipsoid B such that $C_M^Q \cap S \subset B \subset C_M^Q$. This proves the second implication and completes the proof of the theorem.

COROLLARY 4.6.1.

An ALE MY is the Q-minimax estimator for only one ellipsoid (centred at the origin) if and only if there exist a n.n.d. symmetric matrix G such that $MY = (T+G)^{-1}X'V^{-1}Y$.

PROOF. The corollary follows from Theorem 2.4 and the last theorem.

COROLLARY 4.6.2

(i) Each ridge estimator is Q-minimax for only one ellipsoid (centred at the origin).

(ii) Each estimator of the form $t\hat{\beta}$, with $0 < t < 1$, is the Q-minimax estimator for only one ellipsoid (centred at the origin).

PROOF. (i) is a consequence of Corollary 4.6.1 and the form of the ridge estimator. (ii) follows from Corollary 4.6.1 and that $t\hat{\beta} = (T + (\frac{1}{t} - 1)T)^{-1}X'V^{-1}Y$.

The previous results will be applied to some particular estimators which are widely used in practice.

EXAMPLE 4.4.

$M_aY = (T+aI)^{-1}X'V^{-1}Y$, $a > 0$. The risk is given by $Q = I$ (mean square error). Corollary 4.4.1 is useful in this case. We have

$$G = aI, \quad G^{-1} = \frac{1}{a} I,$$

$$B_a^I = (T(T+aI)^{-1} - I)((T+aI)^{-1} - I).$$

$$Tr[B_a^I G^{-1}] = \frac{1}{a} Tr[B_a^I].$$

And

$$R_I[M_a] = Tr[(T + aI)^{-1}] \quad .$$

In most applications V has the form $V = \sigma^2\Gamma$, where σ^2 is "unknown" and Γ is "known". Let $T_0 = X'\Gamma^{-1}X$ and $G = \frac{k}{\sigma^2} I$. Then $(T+G)^{-1} = \sigma^2(T_0+kI)^{-1}$ and $M_kY = (T_0+kI)^{-1}X'\Gamma^{-1}Y$, which is the usual form of ridge estimators. Then

$$B_k^I = (T_0(T_0+kI)^{-1} - I)((T_0+kI)^{-1}T_0-I).$$

$$Tr[B_k^I G^{-1}] = \frac{\sigma^2}{k} Tr[B_k^I] .$$

And

$$R_I[M_k] = \sigma^2 Tr[(T_0 + kI)^{-1}] .$$

Some insight can be obtained if the last results are represented using a set $\{v_i\}_{i=1}^p$ of eigenvectors of T_0 and their eigenvalues λ_i. We will have

$$B_k^I = \sum_{i=1}^p \left(\frac{k}{\lambda_i+k}\right)^2 v_iv_i' ,$$

$$Tr[B_k^I G^{-1}] = \frac{\sigma^2}{k} \sum_{i=1}^p \left(\frac{k}{\lambda_i+k}\right)^2 = \sigma^2 \sum_{i=1}^p \frac{k}{(\lambda_i+k)^2} ,$$

and

$$R_I[M_k] = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i+k} .$$

EXAMPLE 4.5.

Using the same notation as in Example 4.4, let $M_{\tilde{k}}Y = \sum_{i=1}^p \frac{1}{\lambda_i+k_i} v_iv_i' X'V^{-1}Y$, (estimators of this kind are often called generalised ridge estimators). The risk as above will be given by $Q = I$. It is easily realized that all the formulae obtained above in terms of eigenvectors and eigenvalues of T_0 remain valid for $M_{\tilde{k}}Y$ if we replace in them k by k_i, that is

$$B_{\underset{\sim}{k}}^{I} = \sum_{i=1}^{p} \left(\frac{k_i}{\lambda_i+k_i}\right)^2 v_i v_i' ,$$

$$Tr[B_{\underset{\sim}{k}}^{I} G^{-1}] = \sigma^2 \sum_{i=1}^{p} \frac{k_i}{(\lambda_i+k_i)^2} ,$$

and

$$R_I[M_{\underset{\sim}{k}}] = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i+k_i} .$$

EXAMPLE 4.6.

We will study the estimators given in the two previous examples but with the risk given by $Q = T_0$ (prediction risk). We again choose the representation of M_k and $M_{\underset{\sim}{k}}$ in terms of eigenvectors and eigenvalues of T_0. We then have

$$B_k^{T_0} = (T_0(T_0+kI)^{-1} - I)\, T_0((T_0+kI)^{-1}T_0 - I),$$

$$= \left(\sum_{i=1}^{p} \frac{k}{\lambda_i+k} v_i v_i'\right) T_0 \left(\sum_{i=1}^{p} \frac{k}{\lambda_i+k} v_i v_i'\right),$$

$$= \sum_{i=1}^{p} \frac{k^2\lambda_i}{(\lambda_i+k)^2} v_i v_i' .$$

$$Tr[B_k^{T_0} . G^{-1}] = \sigma^2 \sum_{i=1}^{p} \frac{k\lambda_i}{(\lambda_i+k)^2} .$$

And

$$R_{T_0}[M_k] = Tr[(T+G)^{-1}T_0] = \sigma^2\, Tr[(T_0+kI)^{-1}T_0],$$

$$= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i+k} .$$

Again as in the preceding example, the formulae for $M_{\underset{\sim}{k}}Y$ are obtained from the above ones simply by substitution of k by k_i in the right hand side of the formulae and k by $\underset{\sim}{k}$ in the left hand side of the formulae.

EXAMPLE 4.7.

Another very widely used ALE is $M_tY = t\hat{\beta}$, where $0 < t < 1$. We will study first this estimator for the risk given by $Q = I$. We have that $M_tY = t\hat{\beta} = (T + (\frac{1}{t} - 1)T)^{-1}X'V^{-1}Y$ and then $G = (\frac{1}{t} - 1)T$ is a p.d. matrix because $0 < t < 1$. We can apply Corollary 4.4.1. Since $(T+G)^{-1} = tT^{-1}$ we have

$$B_t^I = (TT^{-1}t-I)(tT^{-1}T-I) = (1-t)^2\ I.$$

$$G^{-1} = \frac{t}{1-t}\, T^{-1} \quad .$$

$$Tr[B_t^I\ \ G^{-1}] = (1-t)^2\ \frac{t}{1-t}\ Tr[T^{-1}] = (1-t)t\ Tr[T^{-1}].$$

And

$$R_I[M_t] = Tr[(T+G)^{-1}] = t.Tr[T^{-1}] = \sigma^2.t.Tr[T_0^{-1}] \quad .$$

EXAMPLE 4.7.'

The same estimator as in Example 4.7 but with the risk given by $Q = T_0$. We have

$$B_t^{T_0} = (1-t)^2\ T_0 \ .$$

$$Tr[B_t^{T_0}\ \ G^{-1}] = (1-t)t\ Tr[T_0T^{-1}] = \sigma^2(1-t)tp.$$

And

$$R_{T_0}[M_t] = Tr[(T+G)^{-1}T_0] = \sigma^2 tp \ .$$

(p is the number of dimensions of the parameter β).

EXAMPLE 4.8.

The example we are going to consider is a particular case of all the previous examples of this section. We will suppose that $T = \frac{1}{\sigma^2}\ I$. Then we have $T_0 = I$ and then there is no difference between the mean square error risk and the prediction risk. Also in this case

$M_k Y = M_t Y$ if $\frac{1}{t} - 1 = k$. For this case we have

$$B_t^I = (1-t)^2 I .$$

$$Tr[B_t^I G^{-1}] = \sigma^2(1-t)tp.$$

$$R_I[M_t] = \sigma^2 tp.$$

A case where Example 4.8 applies is when the mean β of a vector Y is to be estimated and Y is assumed to be distributed as $N(\beta,\sigma^2 I_p)$ (N accounts for normal and I_p is the identity in $\mathbb{R}^p$). In this case we have

$$Y = \beta + \varepsilon \quad , \qquad (\text{i.e. } X = I).$$

$$Var[\varepsilon] = \sigma^2 I = V .$$

$$T = X'V^{-1}X = \frac{I}{\sigma^2} \quad , \qquad T_0 = I.$$

And the GLSE $\hat{\beta}$ is $\hat{\beta} = Y$.

Although it is not in the scope of this thesis to treat this kind of problems we will make a small digression just to point to some connections between this work and Empirical Bayes Rules. With this purpose we will enounce a theorem from Baranchik (1964) in our terminology.

BARANCHIK THEOREM.

Let $Y = \beta + \varepsilon$, $\varepsilon \sim N(0,\sigma^2 I)$, $\beta \in \mathbb{R}^p$, $p \geq 3$. Let $\|Y\|^2 = s$. Let f(s) a function of s with values in $\mathbb{R}$. If f(s) is a nondecreasing function of s, $\lim_{s\to\infty} f(s) = \alpha$ and $0 < \alpha \leq 2$ the estimator

$$MY = (1 - \sigma^2(p-2)\frac{f(s)}{s})\, Y,$$

satisfies

$$R_I[M,\beta] \leq \mathrm{Var}_I[\hat{\beta}] \quad , \quad \text{for all } \beta \in \mathbb{R}^p .$$

Since in this case $\hat{\beta} = Y$, MY can be put in the form

$$MY = t(\hat{\beta})\hat{\beta} ,$$

where $t(\hat{\beta})$ is a shrinkage factor which depends on $\hat{\beta}$.

When in estimation problems there is available a good global estimator and a set of estimators which locally are better than the global estimator, one is tempted, instead of using only the global estimator to estimate the value of the unknown parameter, to refine the inference procedure as follows:

(i) First obtain, using the global estimator, an estimate of the region where the unknown value of the parameter lies.
(ii) Then use to estimate the unknown value of the parameter an estimator which is better than the global estimator in the region estimated (by the global estimator).

In our context this procedure would correspond to

(i) Estimate a region where β lies using the GLSE $\hat{\beta}$.
(ii) Estimate β with the Q-minimax estimator corresponding to the region estimated by $\hat{\beta}$.

It is reasonable to think that in those circumstances the region estimated by $\hat{\beta}$ should contain $\hat{\beta}$. This it will be seen does not necessarily hold. From Example 4.8 the ellipsoid where $t\hat{\beta}$ is I-minimax is given by

$$C_t = \{\beta \,|\, (1-t)^2\beta' I\beta \leq \sigma^2(1-t)tp\},$$

or

$$C_t = \{\beta \,|\, \|\beta\|^2 \leq \sigma^2 \frac{t}{1-t} p\} \quad . \tag{4.2.5}$$

Therefore it is a sphere of radius $\sigma(\frac{t}{1-t}\, p)^{\frac{1}{2}}$ and center the origin. In the Baranchik's Theorem we have

$$t(s) = (1 - \sigma^2(p-2)\,\frac{f(s)}{s}\,)\ .$$

Hence

$$\sigma^2\,\frac{t(s)}{1-t(s)}\,p = s(1 - \sigma^2(p-2)\,\frac{f(s)}{s})\,\frac{p}{f(s)(p-2)}\ . \qquad (4.2.6)$$

Then $t(s)\hat{\beta}$ is the I-minimax estimator for the sphere

$$\{\beta \,|\, \|\beta\|^2 \le s(1 - \sigma^2(p-2)\,\frac{f(s)}{s})\frac{p}{f(s)(p-2)}\,\}\ .$$

Therefore the "Baranchik rule" could be interpreted as follows:

(i) The estimated region where β lies is the sphere centred in the origin and of radius given by (4.2.6).

(ii) Use as estimator the I-minimax estimator corresponding to this sphere. Baranchik's Theorem says that (under the hypothesis of the theorem) this procedure is better in the mean square error sense (I-risk) than only to estimate β using the GLSE $\hat{\beta}$.

Now we will use Baranchik's Theorem to see what estimates of the region where β lies will give us a better procedure than the GLSE $\hat{\beta}$ (under the assumptions of the theorem).

Let $C(0,a\|\hat{\beta}\|)$ denote the sphere of center zero and radius $a\|\hat{\beta}\|$. Then from (4.2.5) and (4.2.6) we have

$$a^2 = (1 - \sigma^2(p-2)\,\frac{f(s)}{s})\,\frac{p}{f(s)(p-2)}\ .$$

A few calculations show that

$$f(s) = \frac{p}{p-2}\cdot\frac{s}{a^2 s + \sigma^2 p}\ . \qquad (4.2.7)$$

Then for a fixed a, this function is nondecreasing in s and

$$\lim_{s\to\infty} f(s) = \frac{p}{p-2} \cdot \frac{1}{a^2} .$$

Then f defined as above satisfies Baranchik's Theorem if and only if

$$0 < \frac{p}{p-2} \cdot \frac{1}{a^2} \leq 2 \quad \text{or} \quad \frac{p}{2(p-2)} \leq a^2 .$$

Then Baranchik's Theorem implies (under the hypothesis of the theorem) that the two step procedure proposed will be better than the GLSE if one takes as the estimate of the region where β lies any sphere $C(0, a\|\hat{\beta}\|)$ with $\frac{p}{2(p-2)} \leq a^2$. In particular we have that any sphere centred at the origin which contains Y is a good estimate. We also have that to be a good estimate it does not need to contain Y, because a can be less than 1. If f(s) is substituted in (4.2.6) by the expression in (4.2.7) a simple calculation shows that the shrinkage factor corresponding to a particular value a is

$$t(s) = \frac{a^2 s}{a^2 s + \sigma^2 p} .$$

We will end the section by giving a theorem which provides the tools to calculate the Q-minimax estimators for a wide class of ellipsoids and some examples where it can be applied.

THEOREM 4.7.

Let π a n.n.d. symmetric matrix and B_π its unit ball. Let $T^{-1}\pi T^{-1} = HQH$, where H is a n.n.d. symmetric matrix. Then if

$$A = T^{-1} - \frac{Tr[HQ]}{1 + Tr[\pi T^{-1}]} H \geq 0, \qquad (4.2.8)$$

the Q-minimax estimator of B_π is given by

$$MY = AX'V^{-1}Y .$$

PROOF. From Lemma 4.1 we know that if (4.2.8) holds, MY is an ALE. The Q-bias matrix of MY is

$$B_M^Q = T(A-T^{-1})Q(A-T^{-1})T = \left(\frac{Tr[HQ]}{1 + Tr[\pi T^{-1}]}\right)^2 \pi \quad .$$

Now following the steps of Lemma 4.1 backwards we see that

$$\left(\frac{Tr[HQ]}{1 + Tr[\pi T^{-1}]} \right)^2 = \sum_{i=1}^{p} (1-\delta_i)\delta_i \| \beta_i \|_Q^2 \quad ,$$

where $A = \sum_{i=1}^{p} \delta_i \beta_i \beta_i'$. Theorem 4.5 allows us to conclude that MY is the Q-m.e. for B_π.

The theorem is useful in the sense that it reduces the problem of finding Q-m.e.'s for a wide class of ellipsoids to one of finding Q-eigenvectors and Q-eigenvalues of a matrix; this is because H can easily be obtained from the representation of $T\pi^{-1}T$ in term of Q-eigenvectors. It seems difficult to reduce the problem to a Q-eigenvector problem when the ellipsoid does not satisfy (4.2.8). The next examples are an application of the theorem.

EXAMPLE 4.9.

Let $\pi = \frac{1}{t} I$ and the risk given by Q = I. Then

$$T^{-1}\pi T^{-1} = \frac{1}{t} T^{-2} \;, \quad H = \frac{1}{t^{\frac{1}{2}}} \; T^{-1} \quad ,$$

and

$$M_\pi Y = \frac{t}{t + Tr[T^{-1}]} \; \hat{\beta}.$$

(This result has already been obtained by Lauter, 1975). This estimator is the same as the one in Example 4.1.

EXAMPLE 4.10.

Let $\pi = \frac{1}{t} I$ and the risk given by $Q = T_0$. Then

$$T^{-1} \pi T^{-1} = \frac{1}{t} T^{-2}, \quad H = \frac{1}{\sigma t^{\frac{1}{2}}} T^{-3/2} \quad ,$$

and

$$M_\pi Y = [I - \sigma \frac{Tr[T^{-\frac{1}{2}}]}{t + Tr[T^{-1}]} \quad . \; T^{-\frac{1}{2}}] \; \hat{\beta} \; .$$

EXAMPLE 4.11.

Let $\pi = \frac{1}{t} T$ and the risk given by $Q = I$. Then

$$T^{-1} \pi T^{-1} = \frac{1}{t} T^{-3}, \quad H = \frac{1}{t^{\frac{1}{2}}} \; T^{-3/2} \quad ,$$

and

$$M_\pi Y = [I - \frac{Tr[T^{-3/2}]}{t + p} \; . \; T^{-\frac{1}{2}}] \; \hat{\beta}$$

EXAMPLE 4.12.

Let $\pi = \frac{1}{t} T$ and the risk given by $Q = T_0$. Then

$$T^{-1} \pi T^{-1} = \frac{1}{t} T^{-1}, \quad H = \frac{1}{\sigma t^{\frac{1}{2}}} T^{-1} \quad ,$$

and

$$M_\pi Y = [1 - \frac{p}{t+p}] \hat{\beta} = [\frac{t}{t+p}] \; \hat{\beta} \; .$$

CHAPTER 5

This chapter studies different properties of ALE's. It has three sections:

The first section extends results from Hoerl and Kennard (1970), Farebrother (1975) and Obenchain (1978) among others. The region where an ALE MY has a lower matrix quadratic risk than the GLSE $\hat{\beta}$ is given; this region is an ellipsoid. In Theorem 5.1 and Corollary 5.1.1 different representations and properties of this ellipsoid are given. Theorem 5.2 extends the results of Theorem 5.1 to the non full rank case.

The second section generalizes an optimal property of principal component estimators proved by Fomby et al (1978). It is seen that any S-BLUE enjoys this optimal property if the appropriate Q-quadratic risk is used (Theorem 5.3). This leads to a generalization of the idea of the Marquardt's estimators.

The third section studies a property, proved by Kuks et al (1972) for some ALE's. Theorem 5.4 shows that this property characterizes ALE's. Theorem 5.5 explores further in some aspects of this property.

5.1 WHERE IS AN ALE BETTER THAN THE GLSE?

The section is divided in two parts. The first part concentrates on the study of the problem for the case when X is a full rank matrix; the main result is given in Theorem 5.1 and then it is shown how this result is connected to previous work. The second part generalizes Theorem 5.1 to the case when X is not full rank. We will start giving some definitions.

DEFINITION 5.1.

Let $MY \in L$. The matrix quadratic risk of MY at β, $R[M,\beta]$ is given by

$$R[M,\beta] = E_\beta[(M-\beta)(M-\beta)'] .$$

DEFINITION 5.2.

Let $M_1Y, M_2Y \in L$. We will say that M_1Y is better than M_2Y at β if and only if $R[M_1,\beta] < R[M_2,\beta]$.

When for all $\beta \in W \subset \mathbb{R}^p$, M_1Y is better than M_2Y at β we will say that M_1Y is better than M_2Y for the region W.

When $R[M_1,\beta] \leq R[M_2,\beta]$ we will say that M_1Y is as good as M_2Y at β. M_1Y is as good as M_2Y for the region W will have the obvious meaning.

Since the work of Hoerl and Kennard (1970) conditions which define the region where a ridge estimator is better than the GLSE have been studied. For the estimator $M_kY = (X'X + kI)^{-1}X'Y$ Theobald (1974) gave the following condition: if $k < \frac{2\sigma^2}{\beta'\beta}$ then M_kY is better than the GLSE at β. (Here $V = \sigma^2 I$). In 1976, Farebrother characterized the region W_k where the estimator M_kY is better than the GLSE as follows

$$W_k = \{\beta \mid \beta' U[\tfrac{2}{k} I + \Lambda^{-1}]^{-1} U'\beta < \sigma^2\} ,$$

where $X'X = U'\Lambda U$; Λ is diagonal and U orthogonal.

Let the generalized ridge estimator $M_{\delta} Y = \sum_{i=1}^{p} \frac{\delta_i}{\lambda_i} v_i v_i' X'Y$ (where λ_i are the eigenvalues of X'X and v_i are "their" corresponding eigenvectors). In 1978, Obenchain defined the <u>ridge function of $M_{\delta}Y$</u>

$$RF(\delta) = \sum_{i=1}^{p} \phi_i^2 \frac{1-\delta_i}{1+\delta_i} ,$$

where $\delta = (\delta_1, \ldots \delta_p)$, $\phi_i^2 = \frac{a_i^2 \lambda_i}{\sigma^2}$ and $\beta = \sum_{i=1}^{p} a_i v_i$.

His main result could be read as follows:

Let $\delta = (\delta_1, \ldots, \delta_p)$ such that $0 \leq \delta_i < 1$, i = 1 ... p. Then

(i) $M_{\delta}Y$ is better than the GLSE at β if and only if $RF(\delta) < 1$.

(ii) M_{δ} is as good as the GLSE at β if and only if $RF(\delta) \leq 1$. Kawai and Okamoto (1979) extended those results to the case $0 \leq \delta_i \leq 1$, $1 \leq i \leq p$.

We have then that Theobald, Obenchain, Kawai and Okamoto have answered the following question: given a value β of the parameter what are the conditions on the shrinkage factors of a ridge estimator or a generalized ridge estimator under which the estimator is better than the GLSE at β? Farebrother instead has answered the question: given a ridge estimator $M_k Y$ what are the conditions on β under which $M_k Y$ is better than the GLSE at β? The first point is that those two questions have the same answer, in the sense that the expressions given as answers can be interpreted in two ways: (i) if one considers β fixed then they are conditions on k or δ. (ii) if one considers k or δ fixed then they are conditions on β. We will choose, to express our results, to define conditions on β given MY. The next definition will be useful.

DEFINITION 5.3.

Let $MY \in L$, the good region W_M of MY will be

$$W_M = \{\beta | R[M,\beta] \leq R[\hat{\beta}] \}.$$

Here $R[\hat{\beta}]$ is the matrix quadratic risk of the GLSE and it does not depend on β.

The second point is that ridge or generalized ridge estimators are ALE. It is then natural to ask what is the good region for an arbitrary ALE MY? The next theorem and corollaries will give an answer to this question. A generalization of the expression obtained by Farebrother (1976) and its equivalence to a generalization of the ridge function of Obenchain (1978) is given.

THEOREM 5.1.

Let X be full rank and $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ an ALE.

(a) The good region of MY is

$$W_M = \{\beta | \beta' T(\sum_{i=1}^{p} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i')T\beta \leq 1\} .$$

(b) If $0 < \delta_i < 1$, $i = 1,\ldots,p$; let G such that $MY = (T+G)^{-1}X'V^{-1}Y$, then

$$W_M = \{\beta | \beta'(2G^{-1}+T^{-1})^{-1}\beta \leq 1\}.$$

PROOF. We first will consider that $0 < \delta_i < 1$, $1 \leq i \leq p$. From Theorem 2.4 we have $MY = (T+G)^{-1}X'V^{-1}Y$ with $G = T(\frac{1-\delta_i}{\delta_i} \beta_i \beta_i')T$ a p.d. matrix. Let $A = (T+G)^{-I}$, we have that

$$R[M,\beta] = (AT-I)\beta\beta'(TA-I) + ATA,$$

and

$$R\,[\hat{\beta},\beta] \;=\; T^{-1}.$$

Therefore

$$R[M,\beta] \leq R[\hat{\beta},\beta] \quad , \tag{5.1.1}$$

if and only if

$$(AT-I)\beta\beta'(TA-I) + ATA \leq T^{-1}.$$

Let $D = T^{-1}-ATA$; since $0 < \delta_i < 1$, $1 \leq i \leq p$, D is a p.d. matrix. Then (5.1.1) is equivalent to

$$\alpha' D^{-1}(AT-I)\beta\beta'(TA-I)D^{-1}\alpha \leq \alpha' D^{-1}DD^{-1}\alpha \;, \; \forall\, \alpha \in \mathbb{R}^p .$$

Let $\gamma = (AT-I)\beta$, then the last condition is equivalent to asking

$$(\gamma' D^{-1}\alpha)^2 \leq 1, \quad \forall\, \alpha \in \mathbb{R}^p \text{ such that } \alpha' D^{-1}\alpha = 1.$$

This happens if and only if

$$\gamma' D^{-1}\gamma \leq 1.$$

Now if we notice that $D^{-1} = A^{-1}(A^{-1}T^{-1}A^{-1}-T)^{-1}A^{-1}$ and $(TA-I) = (T-A^{-1})A$, the last condition can be written as

$$\beta'(T-A^{-1})(A^{-1}T^{-1}A^{-1}-T)^{-1}(T-A^{-1})\beta \leq 1.$$

Since $A^{-1} = T+G$ we obtain that (5.1.1) if and only if

$$\beta'(2G^{-1}+T^{-1})^{-1}\beta \leq 1. \tag{5.1.2}$$

This proves (b). Now using Lemma 2.1 an alternative form for (5.1.2) can be obtained; we have

$$G^{-1} = \sum_{i=1}^{p} \frac{\delta_i}{1-\delta_i} \beta_i\beta_i' .$$

$$2G^{-1} + T^{-1} = \sum_{i=1}^{p} (2 \frac{\delta_i}{1-\delta_i} + 1)\beta_i\beta_i' = \sum_{i=1}^{p} \frac{\delta_i+1}{1-\delta_i} \beta_i\beta_i' .$$

And

$$(2G^{-1}+T^{-1})^{-1} = T(\sum_{i=1}^{p} \frac{1-\delta_i}{1+\delta_i} \beta_i\beta_i')T.$$

Then (5.1.1) is equivalent to

$$\beta'T(\sum_{i=1}^{p} \frac{1-\delta_i}{1+\delta_i} \beta_i\beta_i')T\beta \leq 1. \qquad (5.1.3)$$

By continuity this last inequality remains true even if the restriction on δ_i is dropped and we allow $0 \leq \delta_i \leq 1$, $i = 1,\ldots,p$. This proves (a) and the theorem is proved.

The result in (a) tells us that the good region for an ALE

$$MY = \sum_{i=1}^{p} \delta_i\beta_i\beta_i'X'V^{-1}Y$$

is an ellipsoid whose i^{th} T-main axis is in the direction of β_i and has T-length equal to $2(\frac{1+\delta_i}{1-\delta_i})^{\frac{1}{2}}$ (if $\delta_i = 1$, this implies that the ellipsoid is in fact a cylinder; the main i^{th} T-axes is the subspace spanned by β_i).

In most practical situations $V = \sigma^2\Gamma$ (where σ^2 is "unknown" and Γ "known") we will take this into account to change the expressions given in (a) and (b) of the theorem and relate them to previous results and notation.

$$(T+G)^{-1} = (X'\Gamma^{-1}X + \sigma^2 G)^{-1}\sigma^2 .$$

Making $G = \frac{k}{\sigma^2} I$ we obtain

$$(T+G)^{-1}X'V^{-1}Y = (X'\Gamma^{-1}X + kI)^{-1}X'\Gamma^{-1}Y.$$

And (b) becomes

$$\beta'(2\frac{\sigma^2}{k} I + (X'\Gamma^{-1}X)^{-1}\sigma^2)^{-1}\beta \leq 1.$$

Or

$$\beta'(\frac{2}{k} I + (X'\Gamma^{-1}X)^{-1})^{-1}\beta \leq \sigma^2.$$

Let U, Λ, γ such that $X'\Gamma^{-1}X = U' \Lambda U$ and $\gamma = U'\beta$ then the last inequality becomes

$$\gamma'(\frac{2}{k} I + \Lambda^{-1})^{-1}\gamma \leq \sigma^2$$

which is the condition (14) of Farebrother (1976).

We will see how (a) implies the main result of Obenchain (1978) and Kawai et al (1979). Let $\beta = \sum_{i=1}^{p} b_i\beta_i$, that is: the components of β in the T-orthonormal basis

$$\{\beta_i\}_{i=1}^{p} \quad \text{are} \quad \{b_i\}_{i=1}^{p} .$$

We have:

$$\beta'T(\sum_{i=1}^{p} \frac{1-\delta_i}{1+\delta_i} \beta_i\beta_i')T\beta \quad ,$$

$$= (\sum_{i=1}^{p} b_i\beta_i')T(\sum_{i=1}^{p} \frac{1-\delta_i}{1+\delta_i} \beta_i\beta_i')T(\sum_{i=1}^{p} b_i\beta_i),$$

$$= \sum_{i=1}^{p} b_i^2 \frac{1-\delta_i}{1+\delta_i} .$$

The last equality follows from the T-orthonormality of $\{\beta_i\}_{i=1}^{p}$. The last expression is the generalization of the Obenchain's ridge function. To see this we have that

$$M_{\underset{\sim}{\delta}} Y = \sum_{i=1}^{p} \frac{\delta_i}{\lambda_i} v_i v_i' X' \Gamma^{-1} Y = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X' V^{-1} Y$$

and then $\beta_i = \frac{\sigma}{\sqrt{\lambda_i}} v_i$ and so we have

$$\beta = \sum_{i=1}^{p} b_i \beta_i = \sum_{i=1}^{p} b_i \frac{\sigma}{\sqrt{\lambda_i}} v_i = \sum_{i=1}^{p} a_i v_i \quad .$$

Therefore $b_i = \frac{\sqrt{\lambda_i}}{\sigma} a_i$ and

$$\sum_{i=1}^{p} b_i^2 \frac{1-\delta_i}{1+\delta_i} = \sum_{i=1}^{p} a_i^2 \frac{\lambda_i}{\sigma^2} \frac{1-\delta_i}{1+\delta_i} = \sum_{i=1}^{p} \phi_i^2 \frac{1-\delta_i}{1+\delta_i} = RF(\underset{\sim}{\delta}) \; .$$

Part (i) of the next corollary was stated by Farebrother (1976).

COROLLARY 5.1.1.

Let $MY = (T+G)^{-1} X' V^{-1} Y$ an ALE and W_M its good region. Let B_T and $B_{\frac{G}{2}}$ the unit balls of T and $\frac{G}{2}$. Then

(i) $B_T \subset W_M$.

(ii) $B_{\frac{G}{2}} \subset W_M$.

PROOF. The contentions follow immediately from the inequalities:

$$(2G^{-1}+T^{-1})^{-1} \le (0 + T^{-1})^{-1} = T,$$

and

$$(2G^{-1} + T^{-1})^{-1} \le (2G^{-1} + 0)^{-1} = \frac{G}{2} \; .$$

We will now discuss the case when X is not full rank. In this case T is not invertible and there does not exist a GLSE unbiased over $\mathbb{R}^p$. From Theorem 3.3, the only restricted GLSE or S-BLUE which exist are those for which $S \cap N(X) = \{0\}$. Among them there are those which are maximal in the sense that $\mathrm{Span}(S \cup N(X)) = \mathbb{R}^p$. The variance-covariance matrices associated with different maximal S-BLUE's are in general different, then the problem as it was stated in the first part of the section does not make sense. We are compelled to make the following definition.

DEFINITION 5.4.

Let $MY \in L$. Let S such that $S \cap N(X) = \{0\}$ and $\mathrm{Span}(S \cup N(X)) = \mathbb{R}^p$. The <u>S-good region of MY</u> will be:

$$W_M^S = \{\beta | R[M,\beta] \leq R[\hat{\beta}^S]\} .$$

From Theorem 3.5 we know that associated to each subspace S which satisfy the conditions of Definition 5.4, there is a set A_S of admissible linear estimators. We will compare the elements of A_S with the S-BLUE.

THEOREM 5.2.

Let $S \subset \mathbb{R}^p$ a subspace such that $S \cap N(X) = \{0\}$ and $\mathrm{Span}(S \cup N(X)) = \mathbb{R}^p$. Let $MY = \sum_{i=1}^{r} \delta_i \beta_i \beta_i' X'V^{-1}Y$ and ALE such that $\{\beta_i\}_{i=1}^r$ is a T_S-orthonormal basis of S. Then

$$W_M^S = \{\beta \in S | \beta' T(\sum_{i=1}^{r} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i')T\beta \leq 1\}.$$

PROOF. Let T_1 a linear transformation from $\mathbb{R}^p$ to $\mathbb{R}^p$ such that the restrictions of T_1 and T to S are equal and $N(T_1) = S^{\perp}$. Let T_1^- the g-inverse of T_1 such that $T_1^- = \mathrm{Var}[\hat{\beta}^S] = R[\hat{\beta}^S]$. Let P the orthogonal projection onto $S^{\perp}$ and $T_1^{\varepsilon} = T_1 + \frac{1}{\varepsilon} P$, $\varepsilon > 0$. Since

$\{\beta_i\}_{i=1}^r$ is a T_S-orthonormal basis of S it is also a T_1^ε-orthonormal basis of S. Then from Theorem 5.1 we have

$$W_M^\varepsilon = \{\beta | R[M,\beta] \leq [T_1^\varepsilon]^{-1}\} ,$$

$$= \{\beta | \beta' T_1^\varepsilon \left(\sum_{i=1}^{p} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i' \right) T_1^\varepsilon \beta \leq 1\} .$$

Here $\{\beta_i\}_{i=1}^p$ is a T_1^ε - orthonormal basis obtained by completing $\{\beta_i\}_{i=1}^r$.

$$\beta' T_1^\varepsilon \left(\sum_{i=1}^{r} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i' \right) T_1^\varepsilon \beta = \beta' T_1 \left(\sum_{i=1}^{r} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i' \right) T_1 \beta, \quad \text{if } \beta \in S .$$

Therefore

$$W_M^\varepsilon \cap S = \{\beta \in S | \beta' T_1 \left(\sum_{i=1}^{r} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i' \right) T_1 \beta \leq 1\} , \ \forall \varepsilon > 0.$$

Since $(T_1^\varepsilon)^{-1}$ tends to T_1^- as ε tends to zero, this implies that

$$W_M^S \cap S = \{\beta \in S | \beta' T_1 \left(\sum_{i=1}^{r} \frac{1-\delta_i}{1+\delta_i} \beta_i \beta_i' \right) T_1 \beta \leq 1\} .$$

Also since

$$E_\beta[(M-\beta)(M-\beta)'] = (MX-I)\beta\beta'(X'M'-I) + MTM' ,$$

if $\beta = v+u$, with $v \in S$ and $u \in S^\perp$ we have that

$$E_\beta[(M-\beta)(M-\beta)'] = (MX-I)vv'(X'M'-I) + MTM'$$
$$-(MX-I)vu'(X'M'-I) - (MX-I)uv'(X'M'-I) + (MX-I)uu'(X'M'-I)$$

Therefore

$$u'E_\beta[(M-\beta)(M-\beta)']u = \|u\|^4,$$

and

$$u'T_1^- u = 0.$$

Then the only way that

$$E_\beta[(M-\beta)(M-\beta)'] \leq T_1^-,$$

is that u = 0 or equivalently $\beta \in S$. This proves that $\dot{W}_M^S \cap S = W_M^S$. By definition of T_1 we have that for any $\beta \in S$, $T_1\beta = T\beta$ and then

$$W_M^S = \{\beta \in S | \beta'T(\sum_{i=1}^{r} \frac{1-\delta_i}{1+\delta_i} \beta_i\beta_i')T\beta \leq 1\}.$$

This proves the theorem.

5.2 AN OPTIMAL PROPERTY OF S-BLUE's AND GENERALIZED MARQUARDT's ESTIMATORS

The aim of this section is to prove an optimal property of S-BLUE's. This property is related to some ideas behind Marquardt estimators and it will provide us with a way of generalizing them. The property in question is a generalization of one proved by Fomby et al in 1978. The proof we will give here besides being general will be more transparent. We will first state their result. Let

$$A_r = \{MY \mid MY \text{ is an S-BLUE and } \dim S = r\}.$$

Let $\lambda_1 \geq \cdots \geq \lambda_p > 0$ the eigenvalues of T. Let

$$M_r Y = \sum_{i=1}^{r} \frac{1}{\lambda_i} v_i v_i' X' V^{-1} Y,$$

the estimator obtained by deleting the p-r components associated with the smallest eigenvalues of T. Then Fomby et al proved that

$$\mathrm{Var}_I[M_r] \leq \mathrm{Var}_I[M] \quad , \quad \forall\, MY \in A_r.$$

The next propositions will be necessary to prove the generalization of the above property.

PROPOSITION 5.1.

Let Q and T be two p.d. symmetric p×p matrices. Let

$$V_r = \{\Gamma \mid \Gamma = \sum_{i=1}^{r} \gamma_i \gamma_i', \quad \{\gamma_i\}_{i=1}^{r} \text{ is a T-orthonormal set}\} .$$

Let $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_p > 0$ and $\{\beta_i\}_{i=1}^p$ a T-orthonormal set of vectors such that $Q = T(\sum_{i=1}^p \lambda_i\beta_i\beta_i')T$. Let $\Gamma_0 = \sum_{i=1}^r \beta_i\beta_i'$. Then

$$\mathrm{Tr}[Q\Gamma_0] \geq \mathrm{Tr}[Q\Gamma] \quad , \qquad \forall\, \Gamma \in \mathcal{V}_r.$$

PROOF. We have

$$\mathrm{Tr}[Q\Gamma] = \mathrm{Tr}[T(\sum_{i=1}^p \lambda_i\beta_i\beta_i')T\Gamma] \quad ,$$

$$= \sum_{i=1}^p \lambda_i\beta_i'T\Gamma T\beta_i = v'\alpha \quad .$$

where $v' = (\lambda_1,\ldots,\lambda_p)$, $\alpha' = (a_1,\ldots,a_p)$ and $a_i = \beta_i'T\Gamma T\beta_i$, $1 \leq i \leq p$. Since $T\Gamma T \leq T$ and $\|\beta_i\|_T = 1$, $1 \leq i \leq p$, we have that $0 \leq a_i \leq 1$, $1 \leq i \leq p$. Also for any $\Gamma \in \mathcal{V}_r$

$$\sum_{i=1}^p a_i = \sum_{i=1}^p \beta_i'T\Gamma T\beta_i = \mathrm{Tr}[T(\sum_{i=1}^p \beta_i\beta_i')T\Gamma] \quad ,$$

$$= \mathrm{Tr}[T\Gamma] = \sum_{i=1}^r \|\gamma_i\|_T^2 = r.$$

Then our problem can be seen as the one of maximizing the linear functional $v'\alpha$, with α restricted by the following constraints

$$0 \leq a_i \leq 1, \qquad 1 \leq i \leq p,$$

and

$$\sum_{i=1}^p a_i = r.$$

It is easily seen that those restrictions define a convex set whose extreme points are the vectors of $\mathbb{R}^p$ with entries 0 or 1 and with r of the entries equal to 1. It is a well known fact that a linear functional attains its maximum in a convex set at one or more of the extreme points of the convex. Since $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p > 0$ it then becomes clear that if α_0 is such that

$$a_i = 1, \quad 1 \leq i \leq r \quad \text{and} \quad a_i = 0, \quad r < i \leq p, \tag{5.2.1}$$

$v'\alpha$ attains its maximum value at α_0 (in the convex set defined by the constraints). For any $\Gamma \in V_r$, such that $\Gamma = \sum_{i=1}^{r} \gamma_i\gamma_i'$ let $S_\Gamma = \text{Span}\{\gamma_i\}_{i=1}^{r}$ and P_Γ^T the T-projection onto S_Γ. Since $T\Gamma T = T\Gamma T\Gamma T = (P_\Gamma^T)'T(P_\Gamma^T)$, we have

$$\beta_i'T\Gamma T\beta_i = \| P_\Gamma^T(\beta_i)\|_T^2, \qquad 1 \leq i \leq p.$$

Therefore

$$a_i = \beta_i'T\Gamma T\beta_i = 1 \text{ if and only if } \beta_i \in S_\Gamma \quad .$$

The last condition implies that the matrices $\Gamma \in V_r$ which satisfy (5.2.1) only those for which $S_\Gamma = \text{Span}\{\beta_i\}_{i=1}^{r}$. $\Gamma_0 = \sum_{i=1}^{r} \beta_i\beta_i'$ satisfies this condition and this proves the proposition.

PROPOSITION 5.2.

With the notation of the previous proposition, let $\Gamma_1 = \sum_{i=p+1-r}^{p} \beta_i\beta_i'$. Then

$$\text{Tr}[Q\Gamma_1] \leq \text{Tr}[Q\Gamma], \quad \forall\ \Gamma \in V_r.$$

PROOF. Let V_{p-r} defined in the obvious way and $\Lambda \in V_{p-r}$. From Lemma 2.1, (f), for any $\Gamma \in V_r$ there exist at least one $\Lambda \in V_{p-r}$ such that

$$\Gamma + \Lambda = T^{-1} .$$

For the same reason, if $\Lambda_0 = \sum_{i=1}^{p-r} \beta_i\beta_i'$ we have

$$\Gamma_1 + \Lambda_0 = T^{-1} \quad .$$

Then from the previous proposition we have

$$\mathrm{Tr}[Q(T^{-1}-\Gamma_1)] = \mathrm{Tr}[Q\Lambda_0] \geq \mathrm{Tr}[Q\Lambda] = \mathrm{Tr}[Q(T^{-1}-\Gamma)], \quad \forall\, \Gamma \in V_r .$$

This implies the proposition.

PROPOSITION 5.3.

Let $\lambda_1 \geq \ldots \geq \lambda_p > 0$ the T-eigenvalues of Q and suppose that m of them are distinct. Let $n_1,\ldots,n_m$ their multiplicities and $v' = (\lambda_1,\ldots\lambda_p)$. Then

(i) There exists a unique $\Gamma_0 \in V_r$ such that

$$\mathrm{Tr}[Q\Gamma_0] \geq \mathrm{Tr}[Q\Gamma] \quad , \qquad \forall\, \Gamma \in V_r$$

if and only if $r = \sum_{i=1}^{k} n_i$ for some k.

(ii) There exists a unique $\Gamma_1 \in V_r$ such that

$$\mathrm{Tr}[Q\Gamma_1] \leq \mathrm{Tr}[Q\Gamma] \ , \qquad \forall\, \Gamma \in V_r,$$

if and only if $r = \sum_{i=1}^{k} n_i$ for some k.

PROOF. Let $J_r^1 = \{i \mid 1 \leq i \leq r\}$, $J_r^0 = \{i \mid 1 \leq i < r, \text{ or } i = r+1\}$ and J_r any subset of $\{1,\ldots,p\}$ with r elements. Let $\alpha'_{J_r} = (a_1,\ldots a_p)$ such that $a_i = 1$ if $i \in J_r$ and $a_i = 0$ if $i \notin J_r$. Then $v'\alpha_{J_r} = \sum_{i\in J_r} \lambda_i$. If for all k, $r \neq \sum_{i=1}^{k} n_i$, we will have that there exists a k such that $\sum_{i=1}^{k-1} n_i < r < \sum_{i=1}^{k} n_i$ and then $\sum_{i=1}^{k-1} n_i < r < r+1 \leq \sum_{i=1}^{k} n_i$, therefore $\lambda_r = \lambda_{r+1}$ and

$$v'\alpha_{J_r^1} = \sum_{i\in J_r^1} \lambda_i = \sum_{i\in J_r^0} \lambda_i = v'\alpha_{J_r^0} .$$

It is obvious that $\Gamma_{J_r^1}$ and $\Gamma_{J_r^0}$ the matrices of V_r corresponding to J_r^1 and J_r^0 are different. Hence there are at least two matrices in V_r which maximize $\mathrm{Tr}[Q\Gamma]$. Suppose now that $r = \sum_{i=1}^{k} n_i$, then $\lambda_i < \lambda_r$

for all $i > r$; this implies that $\sum_{j \in J_r^1} \lambda_j > \sum_{j \in J_r} \lambda_j$ for all $J_r \neq J_r^1$. This implies that there is a unique matrix in V_r which maximizes $Tr[Q\Gamma]$. This proves (i). The proof of (ii) is similar.

THEOREM 5.3.

Let X be a full rank matrix. Let $Q = T(\sum_{i=1}^{p} \lambda_i \beta_i \beta_i')T$ be a matrix such that $\{\beta_i\}_{i=1}^{p}$ is a T-orthonormal basis of $\mathbb{R}^p$ and $\lambda_1 \geq \cdots \geq \lambda_p > 0$. Let

$$A_r = \{MY | MY \text{ is an S-BLUE and } \dim S = r\},$$

$$M_r Y = \sum_{i=p+1-r}^{p} \beta_i \beta_i' X' V^{-1} Y,$$

and

$$M^r Y = \sum_{i=1}^{r} \beta_i \beta_i' X' V^{-1} Y .$$

Then we have

(i) $Var_Q[M_r] \leq Var_Q[M] \leq Var_Q[M^r] \quad , \quad \forall\, MY \in A_r$.

(ii) $Var_Q[M_r] < Var_Q[M] \quad , \quad \forall\, MY \in A_r, \quad MY \neq M_r Y$
if and only if $r = \sum_{i=1}^{k} n_i$ for some k.

(iii) $Var_Q[M] < Var_Q[M^r], \quad \forall\, MY \in A_r,\ MY \neq M^r Y$

if and only if $r = \sum_{i=1}^{k} n_i$ for some k.

(Here the n_i denote the multiplicities of the different T-eigenvalues of Q).

PROOF. From Corollary 2.2.4 we have that

$$Var[M] \in V_r, \qquad \forall\, MY \in A_r .$$

$$\text{Var}[M_r] = \sum_{i=p+1-r}^{p} \beta_i \beta_i' = \Gamma_1 \quad .$$

And

$$\text{Var}[M^r] = \sum_{i=1}^{r} \beta_i \beta_i' = \Gamma_0 .$$

The results of the theorem follow now from Propositions 5.1, 5.2 and 5.3.

The result of Fomby et al, follows from the first inequality of (i) in the theorem, making Q = I and noticing that "smallest eigenvalues of T" means "largest T-eigenvalues of I" and so to delete the components associated with the smallest eigenvalues of T is equivalent to deleting the components associated with the largest T-eigenvalues of I.

The previous results put in a general context the main idea involved in the "Marquardt estimators". Marquardt suggested the use of estimators obtained by deleting principal components associated with the smallest eigenvalues when $T = X'V^{-1}X$ is ill conditioned. This was proved by Fomby et al to be an optimal choice among the restricted least square estimators with a fixed number of independent restrictions for the minimization of $\text{Var}_I[M]$. When $\text{Var}_Q[M]$ is to be minimized the above results show that the estimators obtained by deleting "Q-principal components" associated with the smallest Q-eigenvalues of T are the best choice.

5.3. THE KUKS-OLMAN PROPERTY

This section will deal with the study of a particular property of ALE's. Kuks and Olman (1972) gave the following property for the estimators $M_G = (T+G)^{-1}X'V^{-1}$, with G a p.d. symmetric matrix.

$$\underset{MY \in L}{\text{Min}} \ \sup_{\beta \in B_G} R_{\alpha\alpha'}[M,\beta] = \sup_{\beta \in B_G} R_{\alpha\alpha'}[M_G,\beta] \ , \ \forall \ \alpha \in \mathbb{R}^p .$$

Here B_G is the unit ball of G.

Since $M_G Y$ is an ALE, one question which arises is if this property can be extended to all the ALE's and how far it can be. This will be the subject of the section. We will introduce some definitions and notation necessary for the discussion. X will be assumed to be full rank.

DEFINITION 5.5.

An estimator MY $\in L$ has the KO property for C if and only if

$$\underset{NY \in L}{\text{Min}} \ \sup_{\beta \in C} R_{\alpha\alpha'}[N,\beta] \geq \sup_{\beta \in C} R_{\alpha\alpha'}[M,\beta] \ , \quad \forall \ \alpha \in \mathbb{R}^p . \qquad (5.3.1)$$

An estimator MY $\in L$ has the KO property if there exist at least one set C for which MY has the KO property.

The result of Kuks and Olman says that the estimator $MY = (T+G)^{-1}X'V^{-1}Y$ has the KO property for B_G.

Given a subset Z of $\mathbb{R}^p$, Sym(Z) and cc(Z) are defined in Definition 4.2. Also $\overline{Z}$ will denote the closure of Z in the usual euclidean topology of $\mathbb{R}^p$. Since $R_{\alpha\alpha'}[M,\beta]$ is a continuous convex function of β and $R_{\alpha\alpha'}[M,\beta] = R_{\alpha\alpha'}[M,-\beta]$ we have that

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M,\beta] = \sup_{\beta \in \overline{cc(Sym(Z))}} R_{\alpha\alpha'}[M,\beta] , \ \forall \quad Z \subset \mathbb{R}^p . \qquad (5.3.2)$$

Given a convex set C, the recession cone of C, Rec (C) is defined as

$$\mathrm{Rec}(C) = \{\beta \mid \beta + \gamma \in C, \quad \forall\, \gamma \in C\} .$$

The recession cone of a convex set C is intuitively the set of unbounded directions of C. If C is bounded Rec(C) = {0}. For more information see Rockafellar (1970), page 60. If C_1 and C_2 are two subsets of $\mathbb{R}^p$, $C_1 + C_2$ will be defined as

$$C_1 + C_2 = \{\beta + \gamma \mid \beta \in C_1,\ \gamma \in C_2\} .$$

The next lemma will give two useful properties for the recession cones of some convex subsets of $\mathbb{R}^p$.

LEMMA 5.1.

Let Z be a non empty subset of $\mathbb{R}^p$ such that

$$Z = \overline{cc\ (\mathrm{Sym}\ (Z)\ )}. \qquad (5.3.3)$$

(i) The recession cone of Z is a subspace S_z contained in Z.

(ii) There exists a set C_z such that

$$Z = C_z + S_z.$$

PROOF. Since Z satisfies (5.3.3), $0 \in Z$. Then if $\alpha \in \mathrm{Rec}(Z)$, $\alpha \in Z$ and $n\alpha \in Z$ for all $n \in N$. Again from (5.3.3) we have that $[-n\alpha, n\alpha]$ is contained in Z and so is Span $\{\alpha\}$. This proves (i). To see (ii) let $C_z = \{\gamma \mid \gamma \perp_T S_z,\ \gamma \in Z\}$. For any $\beta \in \mathbb{R}^p$ we have $\beta = v+u$, $u \in S_z$, $v \in S_z^{\perp T}$. To prove (ii) we only need to see that $v \in C_z$ if and only if $\beta \in Z$. It is evident from (i) and the definition of Rec(Z) that if $v \in C_z$, $\beta = v+u \in Z$ for all $u \in S_z$. Let us now suppose that $\beta \in Z$, then from the definition of Rec(Z), $\beta - u \in Z$, and so $v = \beta - u$ belongs to C_z.

The next proposition gives for any ALE MY a set B_M for which MY has the KO property.

PROPOSITION 5.4.

Let $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ and ALE. Let $G = T(\sum_{\delta_i \neq 0} \frac{1-\delta_i}{\delta_i} \beta_i \beta_i')T$,

$B_G = \{\beta | \beta' G \beta \leq 1\}$, $S_M = \text{Span}\{\beta_i | \delta_i \neq 0\}$. Then MY has the KO property for $B_M = B_G \cap S_M$.

PROOF. From the result of Kuks et al and Theorem 2.4 the proposition is true for the estimators MY with $0 < \delta_i < 1$, $1 \leq i \leq p$; because in this case G is p.d., $S_M = \mathbb{R}^p$ and so $B_M = B_G$. We will suppose now that $0 \leq \delta_i \leq 1$, $1 \leq i \leq p$. Let, for $n \in N$

$$G_n = T(\sum_{0<\delta_i<1} \delta_i \beta_i \beta_i')T + T(\sum_{\delta_i=0} n\beta_i \beta_i')T + T(\sum_{\delta_i=1} \frac{1}{n} \beta_i \beta_i')T \quad .$$

$$B_n = \{\beta | \beta' G_n \beta \leq 1\}.$$

And

$$M_n = (T+G_n)^{-1} X'V^{-1} \quad .$$

It is immediate that

$$\lim_{n\to\infty} M_n = M.$$

And

$$\lim_{n\to\infty} B_n = B_M.$$

We will prove, for $\alpha \in \mathbb{R}^p$, that as n tends to infinity

$$\sup_{\beta \in B_n} R_{\alpha\alpha'}[M_n,\beta] \to \sup_{\beta \in B_M} R_{\alpha\alpha'}[M,\beta] \quad . \tag{5.3.4}$$

The first point to notice is that since B_n is a compact set, for any $\alpha \in \mathbb{R}^p$ there exist β_n (which depends on α) such that

$$\sup_{\beta \in B_n} R_{\alpha\alpha'}[M_n,\beta] = R_{\alpha\alpha'}[M_n,\beta_n]$$

It is not hard to see that

$$\sup_{\beta \in B_n - B_M} \ \inf_{\gamma \in B_M} \|\beta - \gamma\| \to 0, \quad \text{as } n \to \infty . \tag{5.3.5}$$

Then for all $\varepsilon > 0$ there exist n_ε and $\gamma_n \in B_M$ such that if $n > n_\varepsilon$

$$|R_{\alpha\alpha'}[M_n,\beta_n] - R_{\alpha\alpha'}[M,\gamma_n]|$$

$$= |R_{\alpha\alpha'}[M_n,\beta_n] - R_{\alpha\alpha'}[M,\beta_n] + R_{\alpha\alpha'}[M,\beta_n] - R_{\alpha\alpha'}[M,\gamma_n]| < \varepsilon.$$

This implies that

$$\overline{\lim_{n\to\infty}} \sup_{\beta \in B_n} R_{\alpha\alpha'}[M_n,\beta] \leq \sup_{\beta \in B_M} R_{\alpha\alpha'}[M,\beta] \quad . \tag{5.3.6}$$

Now suppose that there exist $\gamma \in B_M$, $\omega > 0$, $n_\omega \in N$ such that for all $n > n_\omega$

$$\sup_{\beta \in B_M} R_{\alpha\alpha'}[M,\beta] = R_{\alpha\alpha'}[M,\gamma] > \sup_{\beta \in B_n} R_{\alpha\alpha'}[M_n,\beta] + \omega \quad .$$

But since M is the limit of M_n and B_M is the limit of B_n, there exist a sequence $\{\beta_n\}_{n=1}^{\infty}$, $\beta_n \in B_n$, such that

$$R_{\alpha\alpha'}[M_n,\beta_n] \to R_{\alpha\alpha'}[M,\gamma] \quad , \qquad \text{as } n \to \infty.$$

This implies

$$\underline{\lim_{n\to\infty}} \sup_{\beta \in B_n} R_{\alpha\alpha'}[M_n,\beta] \geq \sup_{\beta \in B_M} R_{\alpha\alpha'}[M,\beta] \quad . \tag{5.3.7}$$

Then (5.3.4) follows from (5.3.6) and (5.3.7). Now we will prove that MY has the KO property for B_M. Suppose the contrary and let $M_0Y \neq MY$ such that

$$\sup_{\beta\in B_M} R_{\alpha\alpha'}[M,\beta] > \sup_{\beta\in B_M} R_{\alpha\alpha'}[M_0,\beta] \quad , \quad \text{for some } \alpha \in \mathbb{R}^p .$$

Then from (5.3.4), for n big enough and some $0 < \varepsilon < 1$, we will have

$$\sup_{\beta\in B_n} R_{\alpha\alpha'}[M_n,\beta] > \sup_{\beta\in B_M} R_{\alpha\alpha'}[M_0,\beta] + \varepsilon \quad . \tag{5.3.8}$$

Also, for n big enough, from (5.3.5) we have that for all $\beta \in B_n$, there exist $u \in B_n - B_M$ and $v \in B_M$ such that $\beta = v+u$ and

$$\| u'(X'M_0'-I)\alpha\| = \| u'\alpha_0\| < \frac{1}{4}\min(\frac{\varepsilon}{k_0}, \varepsilon) \quad .$$

Where

$$k_0 = \sup_{v\in B_M} \| \alpha'(X'M_0'-I)v\| = \sup_{v\in B_M} \| \alpha_0'v\| \quad .$$

Then for any $\beta \in B_n$

$$R_{\alpha\alpha'}[M_0,\beta] = R_{\alpha\alpha'}[M_0,v] + 2(u'\alpha_0\alpha_0'v) + (u'\alpha_0)^2$$

$$< R_{\alpha\alpha'}[M_0, v] + \varepsilon \quad .$$

Therefore, from (5.3.8)

$$\sup_{\beta\in B_n} R_{\alpha\alpha'}[M_n,\beta] > \sup_{v\in B_M} R_{\alpha\alpha'}[M_0,v] + \varepsilon > \sup_{\beta\in B_n} R_{\alpha\alpha'}[M_0,\beta] \quad .$$

This contradicts the fact that M_nY has the KO property for B_n. The proposition is proved.

THEOREM 5.4.

Let $MY \in L$. Let X be a full rank matrix. Then the following statements are equivalent.

(a) MY is an ALE

(b) MY has the KO property.

PROOF. (a) implies (b) is essentially Proposition 5.4. From (5.3.2) there is no loss of generality considering that $Z = \overline{cc(Sym(Z))}$.
To see the other implication let us first assume that Z is bounded.
Let $\gamma_1 \in Z$ such that

$$\sup_{\beta \in Z} \|\beta\|_T = \|\gamma_1\|_T .$$

From Proposition 2.6, the MMSELE for γ_1, $M_{\gamma_1}Y$ is

$$M_{\gamma_1}Y = \delta_1\beta_1\beta_1'X'V^{-1}Y, \qquad \text{with} \quad \beta_1 = \frac{\gamma_1}{\|\gamma_1\|_T} .$$

If $\alpha_1 = T\beta_1$ we have

$$R_{\alpha_1\alpha_1'}[M_{\gamma_1},\beta] = E_\beta[(\beta_1'T(M_{\gamma_1}-\beta))^2],$$

$$= (\beta_1'T(\delta_1\beta_1\beta_1'T\beta - \beta))^2 + \delta_1^2 = ((\beta_1'T\beta)(\delta_1-1))^2 + \delta_1^2.$$

From the definition of γ_1 we have

$$\sup_{\beta \in Z} (\beta_1'T\beta)^2 = (\beta_1'T\gamma_1)^2 .$$

Therefore

$$\sup_{\beta \in Z} E_\beta[(\beta_1'T(M_{\gamma_1}-\beta))^2] = (\beta_1'T\gamma_1)^2(\delta_1-1)^2 + \delta_1^2$$

$$= E_{\gamma_1}[\beta_1'T(M_{\gamma_1}-\gamma_1)] .$$

From Proposition 2.7, (i)

$$E_{\gamma_1}[(\beta_1{}'T(M_{\gamma_1}-\gamma_1))^2] \leq E_{\gamma_1}[(\beta_1{}'T(M-\gamma_1))^2], \quad \forall \quad MY \in L.$$

Therefore for any $MY \in L$, we have

$$\sup_{\beta \in Z} E_\beta[(\beta_1{}'T(M_{\gamma_1}-\beta))^2] \leq \sup_{\beta \in Z} E_\beta[(\beta_1{}'T(M-\beta))^2].$$

From this we see that a necessary condition for MY to satisfy the KO property in the direction α_1 is that $M = M_{\gamma_1} + M_1$ with $\beta_1{}'TM_1 = 0$ (it is necessary by the unicity of M_{γ_1}). Let now S_1 the subspace T-orthogonal to γ_1. The same argument as above can be restricted to $Z \cap S_1$ to prove that M must be of the form $M = M_{\gamma_1}+M_{\gamma_2}+M_2$ with $\beta_1{}'TM_2 = \beta_2{}'TM_2 = 0$ and $\gamma_1 \perp_T \gamma_2$. After p of those steps we conclude that $M = \sum_{i=1}^{p} M_{\gamma_i}$. Theorem 2.3 implies that MY is an ALE. We will now suppose that Z is not bounded, since $Z = \overline{cc(Sym(Z))}$ from Lemma 5.1, (i), $Rec(Z) = S_z$. If MY is not unbiased on S_z, there exist $\alpha \in \mathbb{R}^p$ and $\beta \in S_z$ such that $\alpha'(MX-I)\beta \neq 0$, this implies that $R_{\alpha\alpha'}[M,\beta]$ is not bounded on Z, but since $R_{\alpha\alpha'}[\hat{\beta},\beta]$ is bounded, MY has not the KO property on Z. This implies that $MY = M_0Y + \hat{\beta}^z$ where $\hat{\beta}^z$ is the S_z-BLUE. By similar arguments to the case where Z is bounded, applied to $S_z^{\perp T} \cap Z$, it can be seen that M_0Y is an ALE and the axis of M_0Y are in $S_z^{\perp T}$. Then again from Theorem 2.3 we can conclude that MY is an ALE. This proves the theorem. The next theorem will characterize the subsets of $\mathbb{R}^p$ for which an ALE MY has the KO property.

THEOREM 5.5.

Let X be a full rank matrix and MY an ALE. Then MY has the KO property for Z if and only if $\overline{cc(Sym(Z))} = B_M$ (B_M is defined in Proposition 5.4).

PROOF. From (5.3.2) on Proposition 5.4 if $\overline{cc(Sym(Z))} = B_M$ MY has the KO property for Z. This proves the sufficient condition. To see the necessary condition, let $MY = \sum_{i=1}^{p} \delta_i \beta_i \beta_i' X'V^{-1}Y$ and suppose first that $0 < \delta_i < 1$, $1 \le i \le p$. Let G the p.d. symmetric matrix such that $M = (T+G)^{-1}X'V^{-1}$. From (5.3.2) there is no loss of generality in assuming $\overline{cc(Sym(Z))} = Z$. Suppose that $Z \neq B_M$, then there exist γ on the boundary of $B_M = B_G$ such that $\sup_{\beta \in Z} \|P_\gamma^G(\beta)\| = \lambda\|\gamma\|$. and $\lambda \neq 1$. Here P_γ^G denotes the G-projection onto Span{γ} or equivalently the projection onto Span{γ} along the hyperplane tangent in γ to B_M. If $\gamma = \sum_{i=1}^{p} b_i \beta_i$ and $\alpha = \sum_{i=1}^{p} a_i T\beta_i$ we have

$$R_{\alpha\alpha'}[M,\gamma] = (\sum_{i=1}^{p} a_i b_i (1-\delta_i))^2 + \sum_{i=1}^{p} a_i^2 \delta_i^2 \quad .$$

If α is chosen such that $\alpha'(MX-I) = \gamma' P_\gamma^G$ we have

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M,\beta] = \sup_{\beta \in Z} (\alpha'(MX-I)\beta)^2 + \sum_{i=1}^{p} a_i^2 \delta_i^2 \ ,$$

$$= (\alpha'(MX-I)\lambda\gamma)^2 + \sum_{i=1}^{p} a_i^2 \delta_i^2 ,$$

$$= \lambda^2 (\sum_{i=1}^{p} a_i b_i (1-\delta_i))^2 + \sum_{i=1}^{p} a_i^2 \delta_i^2 = R_{\alpha\alpha'}[M,\lambda\gamma] \quad . \qquad (5.3.9)$$

Let now $M_\mu = \sum_{i=1}^{p} \mu_i \beta_i \beta_i' X'V^{-1}$ and $\mu(1-\delta_i) = 1-\mu_i$, $1 \le i \le p$. We have that

$$\alpha'(M_\mu X-I) = \mu\alpha'(MX-I).$$

Hence as in (5.3.9) it can be seen that

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M_\mu,\beta] = R_{\alpha\alpha'}[M_\mu,\lambda\gamma] \quad , \qquad (5.3.10)$$

$$= \mu^2 (\sum_{i=1}^{p} \lambda a_i b_i (1-\delta_i))^2 + \sum_{i=1}^{p} a_i^2 (1-\mu(1-\delta_i))^2 \ .$$

Let

$$f_\lambda(\mu) \;=\; R_{\alpha\alpha'}[M_\mu, \lambda\gamma] \;,$$

then the minimum of $f_\lambda(\mu)$ is attained at

$$\mu_\lambda = \frac{\sum_{i=1}^{p} a_i^{\,2}(1-\delta_i)}{\lambda^2\left(\sum_{i=1}^{p} a_i b_i (1-\delta_i)\right)^2 + \sum_{i=1}^{p} a_i^{\,2}(1-\delta_i)^2}$$

Now since $M_\mu = M$ for $\mu = 1$ and MY has the KO property for B_M we have that $\mu_1 = 1$. We also have

(a) If $\lambda > 1$, there exist $\varepsilon > 0$ such that $f_\lambda(\mu) < f_\lambda(1)$ if $\mu \in (1-\varepsilon, 1)$.

(b) If $\lambda < 1$, there exist $\varepsilon > 0$ such that $f_\lambda(\mu) < f_\lambda(1)$ if $\mu \in (1, 1+\varepsilon)$.

Using (5.3.10) this can be rewritten as:

If $\lambda > 1$, $\sup_{\beta \in Z} R_{\alpha\alpha'}[M_\mu, \beta] < \sup_{\beta \in Z} R_{\alpha\alpha'}[M, \beta]$ for $\mu \in (1-\varepsilon, 1)$.

If $\lambda < 1$, $\sup_{\beta \in Z} R_{\alpha\alpha'}[M_\mu, \beta] < \sup_{\beta \in Z} R_{\alpha\alpha'}[M, \beta]$ for $\mu \in (1, 1+\varepsilon)$.

This proves that if $\lambda \neq 1$, MY has not the KO property for Z. We then have proved the case $0 < \delta_i < 1$, $1 \leq i \leq p$.

Let us suppose now that $0 \leq \delta_i < 1$, $1 \leq i \leq p$. Let $S_M = \mathrm{Span}\{\beta_i\}_{\delta_i \neq 0}$ and suppose that $Z \subset S_M$, then a similar argument to the one given for the previous case, but restricted to S_M, shows that $Z = B_M$. Then to prove this case we only need to prove that $Z \subset S_M$. Suppose the contrary, then there exist $\beta_j \in S_M^{\perp T}$ such that $\sup_{\beta \in Z} \| P_j^T(\beta) \| \neq 0$ (P_j^T is the T-projection onto $S_j = \mathrm{Span}\{\beta_j\}$). We will show that there exists an estimator $M_1 Y$ which is better than MY in the direction $\alpha' = \beta_j{}'T$. $M_1 Y$ will be defined as

(a) the MMSELE M_vY of v if $\sup_{\beta \in Z} \| P_j^T(\beta) \| = \| v \|$, $v \in S_j$. (5.3.11)

(b) $\hat{\beta}^j$, the S_j-BLUE if $\sup_{\beta \in Z} \| P_j^T(\beta) \| = \infty$. (5.3.12)

We then have

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M_1,\beta] = \sup_{\beta \in P_j^T(Z)} R_{\alpha\alpha'}[M_1,\beta] ,$$

$$= \frac{\| v \|_T^2}{1 + \| v \|_T^2} , \quad \text{if (5.3.11).}$$

$$= 1 \qquad \text{if (5.3.12).}$$

Since $\alpha' M = 0$, we also have

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M,\beta] = \sup_{\beta \in Z} (\alpha'\beta)^2 = \sup_{\beta \in Z} \| P_j^T(\beta) \|_T^2 ,$$

$$= \| v \|_T^2 , \quad \text{if (5.3.11).}$$

$$= \infty , \quad \text{if (5.3.12).}$$

Then for any of the possibilities (5.3.11) and (5.3.12) we have

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M_1,\beta] < \sup_{\beta \in Z} R_{\alpha\alpha'}[M,\beta]$$

But then MY does not satisfy the KO property for Z. This implies $Z \subset S_M$ and proves that $Z = B_M$ if $0 \le \delta_i < 1$, $1 \le i \le p$. We are left finally with the general case $0 \le \delta_i \le 1$, $1 \le i \le p$. Since $B_M = \overline{cc(Sym(B_M))}$, Lemma 5.1 implies that there exist a set C and a subspace S such that $B_M = C+S$. It is easy to see that $S = Span\{\beta_i\}_{\delta_i=1}$. Since for Z we have assumed (5.3.3), we have also from Lemma 5.1 that $Z = C_z+S_z$. A similar argument to the one given for the case $1 \le \delta_i < 0$, $1 \le i \le p$ restricted to C shows that if MY has the KO property for Z, then $C = C_z$. Then to prove the third case we only need to prove that $S = S_z$. Suppose first that $S_z - S \neq \phi$. Let $\gamma \in S_z - S$, then since MY is biased for γ (MY is only unbiased on S), and $n\gamma \in Z$, $n \in N$, there exists $\alpha \in \mathbb{R}^p$ for which $R_{\alpha\alpha'}[M, \cdot]$ is unbounded on Z. But then MY has not the KO property for Z because $\hat{\beta}$ is better than MY in the direction α (in fact $R_{\alpha\alpha'}[\hat{\beta}, \cdot]$ is bounded on $\mathbb{R}^p$). Then $S_z - S = \phi$. Suppose now that $S - S_z \neq \phi$ and let $\gamma \in S - S_z$ such that $\gamma \perp_T S_z$, then $\sup_{\beta \in Z} \| P_\gamma^T(\beta) \| = \| v \| < \infty$, with $v \in \text{Span}\{\gamma\}$.

Let $\alpha' = \frac{\gamma'}{\| \gamma \|_T} T$ and $M_v Y$ the MMSELE of v, then

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M,\beta] = \sup_{\beta \in P_\gamma^T(Z)} R_{\alpha\alpha'}[M,\beta] = \alpha' \, \text{Var}[M]\alpha = 1.$$

(Because MY is unbiased on S and $\gamma \in S$). We also have

$$\sup_{\beta \in Z} R_{\alpha\alpha'}[M_v,\beta] = \sup_{\beta \in P_\gamma^T(Z)} R_{\alpha\alpha'}[M_v,\beta] = \frac{\| v \|_T^2}{1 + \| v \|_T^2} < 1.$$

(See Proposition 2.8, (iii)). This contradicts the fact that MY has the KO property for Z. Then $S - S_z = \phi$. And we have proved that $S = S_z$. This proves the third case and the theorem.

REFERENCES

Aitken, A.C. (1935). On least squares and linear combination of observations. Proc. Roy. Soc. Edinburgh, Sect. A55, 42-48.

Baranchik, A.J. (1964). Multiple regression and estimation of the mean of the multivariate normal distribution. Tech. Report 51, Stat. Dept. Stanford University.

Bibby, J. (1972). Minimum mean square error estimation, ridge regression and some unanswered questions. Proceedings of the 9th European Meeting of Statisticians, Budapest, 107-121.

Bose, R.C. (1944). The fundamental theorem of linear estimation. Abstract. Proceedings of the Thirty First Indian Science Congress, 4, Part III, 2-3.

Brown, P.J. and Goldstein, M. (1978). Prediction with shrinkage estimators. Math. Operationsforsch Statist. Ser. Statistics, Vol. 9, 3-7.

Bunke, O. (1975). "Minimax linear, ridge and shrunken estimators for linear parameters." Math. Operationsforsch Statisticians 6, 697-701.

Cohen, A. (1966). All admissible linear estimators of the mean vector. Ann. Math. Stat. 37, 458-463.

Durbin, J. and Kendall, M.G. (1951). The geometry of estimation. Biometrika, Vol. 38, 150-158.

Eaton, M.L. (1970). Gauss-Markov estimation for linear models: a coordinate-free approach. Ann. Math. Stat. 41, 528-538.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors, an empirical Bayes approach. JASA, 68, 117-130.

Farebrother, R.W. (1975). The minimum mean square error linear estimator and ridge regression. Technometrics, Vol. 17, 127-128.

Farebrother, R.W. (1976). Further results on the mean square error of ridge regression. JRSS, B, 38, 248-250.

Fomby, T.B., Hill, C.R. and Stan, R.J. (1978). An optimal property of principal components in the context of restricted least squares. JASA, Vol. 73, 191-193.

Gauss, C.F. (1855). Methode des moindres carrés, Paris, Mallet-Bachelier.

Gnot, S., Klonecki, W. and Zmyślony, R. (1980). Best unbiased linear estimation, a coordinate free approach. Probability and Mathematical Statistics, Vol. 1, Fasc. 1, 1-13.

Goldstein, M. and Smith, A.F.M. (1974). Ridge-type estimators for regression analysis. JRSS, B, 36, 284-291.

Gotô, M. (1979). Choice of shrinkage factors in the generalized ridge regression. Math. Japonica 24, No. 2, 153-173.

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, Vol. 12, 55-67.

Hoerl, A.E. and Kennard, R.W. (1970b). Ridge regression applications to non-orthogonal problems. Technometrics 12, 69-82.

Hoffmann, K. (1977). Admissibility of linear estimators with respect to restricted parameter sets. Math. Operationsforsch Statistician Ser. Stat., 8, 425-438.

Kawai, N. and Okamoto, M. (1979). A generalization of the ridge function theorem. Math. Japonica 24, 174-178.

Kruskal, W. (1961). The coordinate-free approach to Gauss-Markov estimation and its application to missing and extra observations. Proceedings of the Fourth Berk. Sym. Math. Stat. Prob. 1, 435-451.

Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach, Ann. Math. Stat. 39, 70-75.

Kruskal, W. (1975). The geometry of generalized inverses. JRSS, B, 37, 272-283.

Kuks, J. (1972). A minimax estimator of regression coefficients (in Russian). Izv. Akad. Nauk Eston. SSR 21, 73-78.

Kuks, J. and Olman, V. (1972). A minimax linear estimator of regression coefficients (in Russian). Izv. Akad. Nauk Eston. SSR 21, 66-72.

Lauter, H. (1975). A minimax linear estimator for linear parameters under restrictions in form of inequalities. Math. Operationsforsch Statisticians 6, 689-695.

Lehmann, E.L. and Scheffé, H. (1950). Completeness, similar regions and unbiased estimation - Part I. Sankhyā 10, 305-340.

Marquardt, D.W. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. Technometrics 12, 591-612.

Mayer, L.S. and Willke, T.A. (1973). On biased estimation in linear models. Technometrics 15, 497-508.

Mitra, S.K. and Rao, C.R. (1971). Generalized inverses of matrices and its applications. Wiley, New York.

Obenchain, R.L. (1978). Good and optimal ridge estimators. Ann. Stat. 6, 1111-1121.

Rao, C.R. (1945). Markoff's theorem with linear restrictions on parameters. Sankhyā 7, 16-19.

Rao, C.R. (1946). On the linear combination of observations and the general theory of least squares. Sankhyā 7, 237-256.

Rao, C.R. (1971). Unified theory of linear estimation. Sankhyā Ser. A, 33, 371-394.

Rao, C.R. (1976). Estimation of parameters in a linear model. Ann. Stat. 4, 1023-1037.

Rockafellar, R.T. (1970). Convex analysis. Princeton University Press.

Rolph, J.E. (1976). Choosing shrinkage estimators for regression problems. Commun. Stat. Theor. Method. A5(9), 789-802.

Seely, J. (1970a). Linear spaces and unbiased estimation. Ann. Math. Stat. 41, 1725-1734.

Seely, J. (1970b). Linear spaces and unbiased estimation - application to the mixed linear model. Ann. Math. Stat. 41, 1735-1748.

Seely, J. and Zyskind, G. (1971). Linear spaces and minimum variance unbiased estimation. Ann. Math. Stat. 42, 691-703.

Shinozaki, N. (1975). A study of generalized inverse of matrix and estimation with quadratic loss. Ph.D. Thesis submitted to the Keio University, Japan.

Stein, C.M. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Stat. Prob. 1, 197-206. University of California Press.

Theil, H. (1971). Principles of econometrics. North Holland Publishing Company, Amsterdam.

Thisted, R. (1976). Ridge regression, minimax estimation and empirical Bayes methods. Technical Report No. 28, Division of Biostatistics, Stanford University.

Toutembourg, H. and Bibby, J. (1977). Prediction and improved estimation in linear models. Wiley, Berlin.

Watson, G.S. (1967). Linear least squares regression. Ann. Math. Stat. 38, 1679-99.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best simple least squares estimators in linear models. Ann. Math. Stat., 38, 1092-1109.

Theobald, C.M. (1974). Generalizations of mean square error applied to ridge regression. J.R. Statist. Soc. B, 36, 103-106.